Asymptotic analysis of the role of spatial sampling for hyper-parameter estimation of Gaussian processes

François Bachoc[a,b,*]

[a] *CEA-Saclay, DEN, DM2S, STMF, LGLS, F-91191 Gif-Sur-Yvette, France*
[b] *Laboratoire de Probabilités et Modèles Aléatoires, Université Paris VII Site Chevaleret, case 7012, 75205 Paris cedex 13*

Abstract

Hyper-parameter estimation of Gaussian processes is analyzed in an asymptotic framework. The spatial sampling is a randomly perturbed regular grid and its deviation from the perfect regular grid is controlled by a single scalar regularity parameter. Consistency and asymptotic normality are proved for the Maximum Likelihood and Cross Validation estimators of the hyper-parameters. The asymptotic covariance matrices of the hyper-parameter estimators are deterministic functions of the regularity parameter. By means of an exhaustive study of the asymptotic covariance matrices, it is shown that irregular sampling is generally an advantage to estimation, but we identify cases where it is not the case. Therefore, a negative answer is given to the claim that irregular sampling is always better for hyper-parameter estimation than regular sampling.

Keywords: Uncertainty quantification, metamodel, Kriging, hyper-parameter estimation, maximum likelihood, leave-one-out, increasing-domain asymptotic

1. Introduction

In many areas of science that involve measurements or data acquisition, one often has to answer the question of how the set of experiments should be designed [13]. It is known that in many situations, an irregular, or even random, spatial sampling is preferable to a regular one. Examples of these situations are found in many fields. For numerical integration, Gaussian quadrature rules generally yield irregular grids [16, ch.4]. The best known low-discrepancy sequences for quasi-Monte Carlo methods (van der Corput, Halton, Sobol, Faure, Hammersley,...) are not regular either [14]. In the compressed sensing domain, it has been shown that one can recover a signal very efficiently, and at a small cost, by using random measurements [4].

In this paper, we are focused on the role of spatial sampling for meta-modeling. Meta-modeling is particularly relevant for the analysis of complex computer models [20]. We will address the case of Kriging models, which consist in interpolating the values of a Gaussian random field given observations at a finite set of observation points. Kriging has become a popular method for a large range of applications, such as numerical code approximation [19, 20] and calibration [15] or global optimization [9].

One of the main issues regarding Kriging is the choice of the covariance function for the Gaussian process. Indeed, a Kriging model yields an unbiased predictor with minimal variance and a correct predictive variance only if the correct covariance function is used. The most common practice is to statistically estimate the covariance function, from a set of observations

[☆] A supplementary material is attached in the electronic version of the article.

[*] Corresponding author: François Bachoc
CEA-Saclay, DEN, DM2S, STMF, LGLS, F-91191 Gif-Sur-Yvette, France
Phone: +33 (0) 1 69 08 97 91
Email: francois.bachoc@cea.fr

of the Gaussian process, and to plug [24, ch.6.8] the estimate in the Kriging equations. Usually, it is assumed that the covariance function belongs to a given parametric family (see [1] for a review of classical families). In this case, the estimation boils down to estimating the corresponding parameters, that are called "hyper-parameters".

The spatial sampling, and particularly its degree of regularity, play an important role for the covariance function estimation. In his monograph, Stein shows that a highly irregular sampling, with pairs of very close observation points, is preferable over a regular grid for the estimation of the smoothness parameter of the Matérn model [24, ch.6.9]. It is shown in [32] that the optimal samplings, for maximizing the log of the determinant of the Fisher information matrix, are extremely irregular. Therefore, a generally admitted conjecture is that irregular sampling is always better, for hyper-parameter estimation in Kriging, than regular sampling.

In this paper, we address this conjecture in an asymptotic framework. Since exact finite-sample results are generally not reachable and not meaningful as they are specific to the situation, asymptotic theory is widely used to give approximations of the estimated hyper-parameter distribution.

The two most studied asymptotic frameworks are the increasing-domain and fixed-domain asymptotics [24, p.62]. In increasing-domain asymptotics, a minimal spacing exists between two different observation points, so that the infinite sequence of observation points is unbounded. In fixed-domain asymptotics, the sequence is dense in a bounded domain.

In fixed-domain asymptotics, significant results are available, concerning the estimation of the covariance function, and its influence on Kriging predictions. In this asymptotic framework, two types of covariance hyper-parameters can be distinguished: microergodic and non-microergodic hyper-parameters. Following the definition in [24], an hyper-parameter is microergodic if two covariance functions are orthogonal whenever they differ for it (as in [24], we say that two covariance functions are orthogonal if the two underlying Gaussian measures are orthogonal). Non-microergodic hyper-parameters cannot be consistently estimated, but have no asymptotic influence on Kriging predictions [21, 22, 23, 30]. On the contrary, there is a fair amount of literature on consistently estimating microergodic hyper-parameters using the Maximum Likelihood (ML) method. Consistency has been proved for several models [28, 29, 11, 30, 10, 2]. Micro-ergodic hyper-parameters have an asymptotic influence on predictions, as shown in [27, ch.5].

Nevertheless, the fixed-domain asymptotic framework is not well adapted to study the influence of the irregularity of the spatial sampling on hyper-parameter estimation. Indeed, we would like to compare sampling techniques by inspection of the asymptotic distributions of the hyper-parameter estimators. In fixed-domain asymptotics, when an asymptotic distribution is proved for ML [28, 29, 6], it turns out that it is independent of the dense sequence of observation points. This makes it impossible to compare the effect of spatial sampling on hyper-parameter estimation using fixed-domain asymptotics techniques.

The first characteristic of increasing-domain asymptotics is that, as shown in section 5, all the hyper-parameters have strong asymptotic influences on predictions. The second characteristic is that all the hyper-parameters (satisfying a very general identifiability assumption) can be consistently estimated, and that asymptotic normality generally holds [26, 12, 5]. Roughly speaking, increasing-domain asymptotics is characterized by a vanishing dependence between distant observation points. As a result, a large sample size gives more and more information about the covariance structure. Finally, we show that the asymptotic variances of the hyper-parameter estimators strongly depend on the spatial sampling. This is why we address the increasing-domain asymptotic framework to study the influence of the spatial sampling on the hyper-parameter estimation.

We propose a sequence of random spatial samplings of size $n \in \mathbb{N}^*$. The regularity of the spatial sampling sequence is characterized by a regularity parameter $\epsilon \in (-\frac{1}{2}, \frac{1}{2})$. $\epsilon = 0$ corresponds to a regular grid, and the irregularity is increasing with ϵ. We study the ML estimator, and also a Cross Validation (CV) estimator [25, 31], for which, to the best of our knowledge, no asymptotic results are yet available in the literature. For both estimators, we prove an asymptotic normality result for the estimation, with a $\sqrt{n}$ convergence, and an asymptotic covariance

matrix which is a deterministic function of ϵ. The asymptotic normality yields, classically, approximate confidence intervals for finite-sample estimation. Then, carrying out an exhaustive analysis of the asymptotic variance, for the one-dimensional Matérn model, we show that irregularity is indeed an advantage for estimation in the majority of the cases. However, we can determine the cases in which a regular grid performs better for estimation than irregular ones. This definitively gives a negative answer to the claim that irregular sampling is always better for hyper-parameter estimation than regular sampling.

The rest of the article is organized as follows. In section 2, we introduce the random sequence of observation points, that is parameterized by the regularity parameter ϵ. We also present the ML and CV estimators. In section 3, we give the asymptotic normality results. In section 4, we carry out and exhaustive study of the asymptotic variance. In section Appendix A, we prove the results of section 3, in section Appendix B, we prove the results of section 4, and in section Appendix C, we state and prove several technical results. Finally, section Appendix D is dedicated to the one-dimensional case, with $\epsilon = 0$. We present an efficient calculation of the asymptotic variances for ML and CV and of the second derivative of the asymptotic variance of ML, at $\epsilon = 0$, using properties of Toeplitz matrix sequences.

2. Context

2.1. Presentation and notation for the spatial sampling sequence

We consider a stationary Gaussian process Y on $\mathbb{R}^d$. We denote $\Theta = [\theta_{inf}, \theta_{sup}]^p$. The covariance function of Y is K_{θ_0} with $\theta_{inf} < (\theta_0)_i < \theta_{sup}$, for $1 \leq i \leq p$. K_{θ_0} belongs to a parametric model $\{K_\theta, \theta \in \Theta\}$, with K_θ a stationary covariance function.

We shall assume the following condition for the parametric model, which is satisfied in all the most classical cases, and especially for the Matérn model that we will analyze in detail in section 4.

Condition 2.1.

- *For all $\theta \in \Theta$, the covariance function K_θ is stationary.*
- *The covariance function K_θ is three times differentiable with respect to θ. For all $q \in \{0, ..., 3\}$, $i_1, ..., i_q \in \{1, ..., p\}$, there exists $C_{i_1,...,i_q} < +\infty$ so that for all $\theta \in \Theta$, $t \in \mathbb{R}^d$,*

$$\frac{\partial}{\partial \theta_{i_1}} ... \frac{\partial}{\partial \theta_{i_q}} K_\theta(t) \leq \frac{C_{i_1,...,i_q}}{1 + |t|^{d+1}}, \tag{1}$$

where $|t|$ is the Euclidian norm of t. We define the Fourier transform of a function $h : \mathbb{R}^d \to \mathbb{R}$ by $\hat{h}(f) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} h(t) e^{-\mathrm{i} f \cdot t} dt$, where $\mathrm{i}^2 = -1$. Then, for all $\theta \in \Theta$, the covariance function K_θ has a Fourier transform $\hat{K}_\theta$ that is continuous and bounded.

- *For all $\theta \in \Theta$, K_θ satisfies*

$$K_\theta(t) = \int_{\mathbb{R}^d} \hat{K}_\theta(f) e^{\mathrm{i} f \cdot t} df.$$

- *$(\theta, f) \to \hat{K}_\theta(f)$ is continuous and positive on $\Theta \times \mathbb{R}^d$.*

We denote by $(v_i)_{i \in \mathbb{N}^*}$ a sequence of deterministic points in $\mathbb{N}^d$ so that for all $N \in \mathbb{N}^*$, $\{v_i, 1 \leq i \leq N^d\} = \{1, ..., N\}^d$. Y is observed at the points $v_i + \epsilon X_i$, $1 \leq i \leq n$, $n \in \mathbb{N}^*$, with $-\frac{1}{2} < \epsilon < \frac{1}{2}$ and $X_i \sim_{iid} \mathcal{L}_X$. $\mathcal{L}_X$ is a symmetric probability law with support $S_X \subset [-1, 1]^d$, and with a positive probability density on S_X. Two remarks can be made on this sequence of observation points:

- This is an increasing-domain asymptotic context. The condition $-\frac{1}{2} < \epsilon < \frac{1}{2}$ ensures a minimal spacing between two distinct observation points.

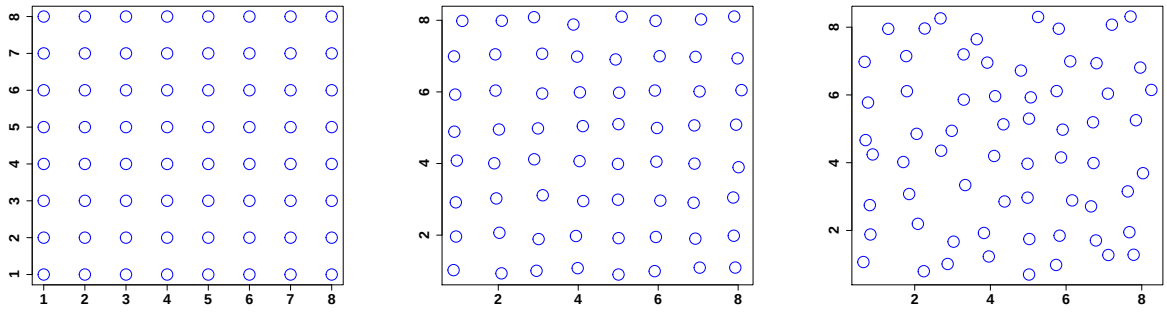

Figure 1: Examples of three perturbed grids. The dimension is $d = 2$ and the number of observation points is $n = 8^2$. From left to right, the values of the regularity parameter are 0, $\frac{1}{8}$ and $\frac{3}{8}$. $\epsilon = 0$ corresponds to a regular observation grid, while, when $|\epsilon|$ is close to $\frac{1}{2}$, the observation set is highly irregular.

- The observation sequence we study is random, and the parameter ϵ is a regularity parameter. $\epsilon = 0$ corresponds to a regular observation grid, while, when $|\epsilon|$ is close to $\frac{1}{2}$, the observation set is highly irregular. Examples of observation sets are given in figure 1, with $d = 2$, $n = 8^2$, and different values of ϵ.

We denote $C_{S_X} = \{t_1 - t_2, t_1 \in S_X, t_2 \in S_X\}$, the set of all possible differences between two points in S_X. We denote, for $n \in \mathbb{N}^*$, $X = (X_1, ..., X_n)$, where we do not write explicitly the dependence in n for clarity. X is a random variable with law $\mathcal{L}_X^{\otimes n}$. We also denote $x = (x_1, ..., x_n)$, an element of $(S_X)^n$, as a realization of X.

We define the $n \times n$ random matrix R_θ by $(R_\theta)_{i,j} = K_\theta(v_i - v_j + \epsilon(X_i - X_j))$. We do not write explicitly the dependence of R_θ with respect to X, ϵ and n. We shall denote, as a simplification, $R := R_{\theta_0}$. We define the random vector y of size n by $y_i = Y(v_i + \epsilon X_i)$. We do not write explicitly the dependence of y with respect to X, ϵ and n.

We denote as in [8], for a real $n \times n$ matrix A, $|A|^2 = \frac{1}{n}\sum_{ij,=1}^n A_{i,j}^2$ and $||A||$ the largest singular value of A. $|.|$ and $||.||$ are norms and $||.||$ is a matrix norm. We denote by $\phi_i(M)$, $1 \leq i \leq n$, the eigenvalues of a symmetric matrix M. We denote, for two sequences of square matrices A and B, depending on $n \in \mathbb{N}^*$, $A \sim B$ if $|A - B| \to_{n\to+\infty} 0$ and $||A||$ and $||B||$ are bounded with respect to n. Finally, for a square matrix A, we denote by $diag(A)$ the matrix obtained by setting to 0 all non diagonal elements of A.

Finally, for a sequence of real random variables z_n, we denote $z_n \to_p 0$ and $z_n = o_p(1)$ when z_n converges to zero in probability.

2.2. ML and CV estimators

We denote $L_\theta := \frac{1}{n}\left\{\log(\det(R_\theta)) + y^t R_\theta^{-1} y\right\}$ the modified opposite log-likelihood, where we do not write explicitly the dependence in X, Y, n and ϵ. We denote by $\hat{\theta}_{ML}$ the Maximum Likelihood estimator, defined by

$$\hat{\theta}_{ML} \in \underset{\theta\in\Theta}{\operatorname{argmin}} L_\theta, \tag{2}$$

where we do not write explicitly the dependence of $\hat{\theta}_{ML}$ with respect to X, Y, ϵ and n.

Remark. The ML estimator in (2) is actually not entirely defined, since the likelihood function of (2) can have more than one global minimizer. Nevertheless, the convergence results of $\hat{\theta}_{ML}$, as $n \to +\infty$, hold when $\hat{\theta}_{ML}$ is any random variable belonging to the set of the global minimizers of the likelihood of (2), regardless of the value chosen in this set. Furthermore, it can be shown that, with probability converging to one, as $n \to \infty$ (see the proof of Proposition Appendix C.10 in Appendix Appendix C), the likelihood function has a unique global minimum. To define a

measurable function $\hat{\theta}_{ML}$ of Y and X, belonging to the set of the minimizers of the likelihood, one possibility is the following. For a given realization of Y and X, let $\mathcal{K}$ be the set of the minimizers of the likelihood. Let $\mathcal{K}_0 = \mathcal{K}$ and, for $0 \leq k \leq p-1$, $\mathcal{K}_{k+1}$ is the subset of $\mathcal{K}_k$ whose elements have their $k+1$th coordinates equal to $\min\left\{\tilde{\theta}_{k+1}, \tilde{\theta} \in \mathcal{K}_k\right\}$. Since, $\mathcal{K}$ is compact (because the likelihood function is continuous with respect to θ and defined on the compact set Θ), the set $\mathcal{K}_p$ is composed of a unique element, that we define as $\hat{\theta}_{ML}$, which is a measurable function of X and Y. The same remark can be made for the Cross Validation estimator of (3).

When the increasing-domain asymptotics sequence of observation points is deterministic, it is shown in [12] that $\hat{\theta}_{ML}$ converges to a centered Gaussian random vector. The asymptotic covariance matrix is the inverse of the Fisher information matrix. For fixed n, the Fisher information matrix is the $p \times p$ matrix with (i,j)th element equal to $\frac{1}{2}\mathrm{Tr}\left(\mathrm{R}_{\theta_0}^{-1}\frac{\partial \mathrm{R}_{\theta_0}}{\partial \theta_\mathrm{i}}\mathrm{R}_{\theta_0}^{-1}\frac{\partial \mathrm{R}_{\theta_0}}{\partial \theta_\mathrm{j}}\right)$. Since the literature has not addressed yet the asymptotic distribution of $\hat{\theta}_{ML}$ in increasing-domain asymptotics with random observation points, we give complete proofs about it in section Appendix A. Our techniques are original and not specifically oriented towards ML contrary to [26, 12, 5], so that they allow us to address the asymptotic distribution of the CV estimator in the same fashion.

The CV estimator is defined by

$$\hat{\theta}_{CV} \in \underset{\theta\in\Theta}{\operatorname{argmin}} \sum_{i=1}^{n} \{y_i - \hat{y}_{i,\theta}\left(y_{-i}\right)\}^2, \tag{3}$$

where, for $1 \leq i \leq n$, $\hat{y}_{i,\theta}\left(y_{-i}\right) := \mathbb{E}_{\theta|X}\left(y_i | y_1, ..., y_{i-1}, y_{i+1}, ..., y_n\right)$ is the Kriging Leave-One-Out prediction of y_i with covariance hyper-parameters θ. $\mathbb{E}_{\theta|X}$ denotes the expectation with respect to the distribution of Y with the covariance function K_θ, given X.

The CV estimator selects the hyper-parameters according to the criterion of the point-wise prediction errors. This criterion does not involve the Kriging predictive variances. Hence, the CV estimator of (3) cannot estimate an hyper-parameter impacting only on the variance of the Gaussian process. Nevertheless, all the classical parametric models $\{K_\theta, \theta \in \Theta\}$ satisfy the decomposition $\theta = \left(\sigma^2, \tilde{\theta}\right)$ and $\{K_\theta, \theta \in \Theta\} = \{\sigma^2 \tilde{K}_{\tilde{\theta}}, \sigma^2 > 0, \tilde{\theta} \in \tilde{\Theta}\}$, with $\tilde{K}_{\tilde{\theta}}$ a correlation function. Hence, in this case, $\tilde{\theta}$ would be estimated by (3), and σ^2 would be estimated by the equation $\hat{\sigma}^2_{CV}(\tilde{\theta}) = \frac{1}{n}\sum_{i=1}^{n} \frac{\{y_i - \hat{y}_{i,\tilde{\theta}}(y_{-i})\}^2}{c^2_{i,-i,\tilde{\theta}}}$, where $c^2_{i,-i,\tilde{\theta}} := var_{\tilde{\theta}|X}\left(y_i | y_1, ..., y_{i-1}, y_{i+1}, ..., y_n\right)$ is the Kriging Leave-One-Out predictive variance for y_i with hyper-parameters $\sigma^2 = 1$ and $\tilde{\theta}$. $var_{\tilde{\theta}|X}$ denotes the variance with respect to the distribution of Y with the covariance function K_θ, $\theta = (1, \tilde{\theta})$, given X. To summarize, the general CV procedure we study is a two-step procedure. In a first step, the correlation hyper-parameters are selected according to a mean square error criterion. In a second step, the global variance hyper-parameter is selected, so that the predictive variances are adapted to the Leave-One-Out prediction errors. Here, we address the first step, so we focus on the CV estimator defined in (3).

The criterion (3) can be computed with a single matrix inversion, by means of virtual LOO formulas (see e.g [18, ch.5.2] for the zero-mean case addressed here, and [7] for the universal Kriging case). These virtual LOO formulas yield

$$\sum_{i=1}^{n} \left\{y_i - \hat{y}_{i,\theta}\left(y_{-i}\right)\right\}^2 = y^t R_\theta^{-1} diag\left(R_\theta^{-1}\right)^{-2} R_\theta^{-1} y,$$

which is useful both in practice (to compute quickly the LOO errors) and in the proofs on CV. We then define

$$CV_\theta := \frac{1}{n} y^t R_\theta^{-1} diag\left(R_\theta^{-1}\right)^{-2} R_\theta^{-1} y$$

as the CV criterion, where we do not write explicitly the dependence in X, n, Y and ϵ. Hence we have, equivalently to (3), $\hat{\theta}_{CV} \in \operatorname{argmin}_{\theta\in\Theta} CV_\theta$.

Since the asymptotic covariance matrix of the ML estimator is the inverse of the Fisher information matrix, this estimator should be used when $\theta_0 \in \Theta$ holds, which is the case of interest here. However, in practice, it is likely that the true covariance function of the Gaussian process does not belong to the parametric family used for the estimation. In [3], this case is called model misspecification case, and it is shown that CV is more efficient that ML in this case. Hence, the CV estimator is relevant in practice, which is a reason for studying it in the well-specified case $\theta_0 \in \Theta$ addressed here. Hence we aim at studying the influence of the spatial sampling on CV as well as on ML. Furthermore, since it is expected that ML performs better than CV in the well-specified case, we are interested in quantifying this fact.

3. Consistency and asymptotic normality

Proposition 3.1 addresses the consistency of the ML estimator. The only assumption on the parametric family of covariance functions is an identifiability assumption. Basically, for a fixed ϵ, there should not exist two distinct hyper-parameters so that the two associated covariance functions are the same, on the set of inter-point distances covered by the random spatial sampling. The identifiability assumption is clearly minimal.

Proposition 3.1. *Assume that condition 2.1 is satisfied.*

For $\epsilon = 0$, if there does not exist $\theta \neq \theta_0$ so that $K_\theta(v) = K_{\theta_0}(v)$ for all $v \in \mathbb{Z}^d$, then the ML estimator is consistent.

For $\epsilon \neq 0$, we denote $D_\epsilon = \cup_{v \in \mathbb{Z}^d \setminus 0} (v + \epsilon C_{S_X})$, with $C_{S_X} = \{t_1 - t_2, t_1 \in S_X, t_2 \in S_X\}$. Then, if there does not exist $\theta \neq \theta_0$ so that $K_\theta = K_{\theta_0}$ a.s. on D_ϵ, according to the Lebesgue measure on D_ϵ, and $K_\theta(0) = K_{\theta_0}(0)$, then the ML estimator is consistent.

In proposition 3.2, we address the asymptotic normality of ML. The convergence rate is $\sqrt{n}$, as in a classical *iid* framework, and we prove the existence of a deterministic asymptotic covariance matrix of $\sqrt{n}\hat{\theta}_{ML}$ which depends only on the regularity parameter ϵ. In proposition 3.3, we prove that this asymptotic covariance matrix is positive, as long as the different derivative functions with respect to θ at θ_0 of the covariance function are non redundant on the set of inter-point distances covered by the random spatial sampling. This condition is minimal, since when these derivatives are redundant, the Fisher information matrix is singular for all finite sample-size n and its kernel is independent of n.

Proposition 3.2. *Assume that condition 2.1 is satisfied.*

For all $1 \leq i, j \leq p$, the random trace $\frac{1}{n}\mathrm{Tr}\left(\mathrm{R}^{-1}\frac{\partial \mathrm{R}}{\partial \theta_\mathrm{i}}\mathrm{R}^{-1}\frac{\partial \mathrm{R}}{\partial \theta_\mathrm{j}}\right)$ converges a.s. to the element $(\Sigma_{ML})_{i,j}$ of a $p \times p$ deterministic matrix Σ_{ML} as $n \to +\infty$.

If $\hat{\theta}_{ML}$ is consistent and if Σ_{ML} is positive, then

$$\sqrt{n}\left(\hat{\theta}_{ML} - \theta_0\right) \to \mathcal{N}\left(0, 2\Sigma_{ML}^{-1}\right).$$

Proposition 3.3. *Assume that condition 2.1 is satisfied.*

For $\epsilon = 0$, if there does not exist $v_\lambda = (\lambda_1, ..., \lambda_p) \in \mathbb{R}^p$, v_λ different from zero, so that $\sum_{k=1}^p \lambda_k \frac{\partial}{\partial \theta_k} K_{\theta_0}(v) = 0$ for all $v \in \mathbb{Z}^d$, then Σ_{ML} is positive.

For $\epsilon \neq 0$, we denote $D_\epsilon = \cup_{v \in \mathbb{Z}^d \setminus 0} (v + \epsilon C_{S_X})$, with $C_{S_X} = \{t_1 - t_2, t_1 \in S_X, t_2 \in S_X\}$. If there does not exist $v_\lambda = (\lambda_1, ..., \lambda_p) \in \mathbb{R}^p$, v_λ different from zero, so that $t \to \sum_{k=1}^p \lambda_k \frac{\partial}{\partial \theta_k} K_{\theta_0}(t)$ is almost surely zero on D_ϵ, with respect to the Lebesgue measure on D_ϵ, and that $\sum_{k=1}^p \lambda_k \frac{\partial}{\partial \theta_k} K_{\theta_0}(0)$ is null, then Σ_{ML} is positive.

Proposition 3.4 addresses the consistency of the CV estimator. The identifiability assumption is required, like in the ML case. Since the CV estimator is designed for estimating correlation hyper-parameters, we assume that the parametric model $\{K_\theta, \theta \in \Theta\}$ contains only correlation functions. This assumption holds in most classical cases, and yields results that are easy to express and interpret. The case of hybrid hyper-parameters, specifying both a variance and a

correlation structure should be consistently estimated by the CV estimator. Nevertheless, since such hyper-parameters do not exist in the classical families of covariance function, we do not address this case.

Proposition 3.4. *Assume that condition 2.1 is satisfied and that for all* $\theta \in \Theta$, $K_\theta(0) = 1$.

For $\epsilon = 0$, *if there does not exist* $\theta \neq \theta_0$ *so that* $K_\theta(v) = K_{\theta_0}(v)$ *for all* $v \in \mathbb{Z}^d$, *then the CV estimator is consistent.*

For $\epsilon \neq 0$, *we denote* $D_\epsilon = \cup_{v \in \mathbb{Z}^d \setminus 0} (v + \epsilon C_{S_X})$, *with* $C_{S_X} = \{t_1 - t_2, t_1 \in S_X, t_2 \in S_X\}$. *Then, if their does not exist* $\theta \neq \theta_0$ *so that* $K_\theta = K_{\theta_0}$ *a.s. on* D_ϵ, *with respect to the Lebesgue measure on* D_ϵ, *the CV estimator is consistent.*

Proposition 3.5 gives the expression of the covariance matrix of the gradient of the CV criterion CV_θ and of the mean matrix of its Hessian. These moments are classically used in statistics to prove asymptotic distributions of consistent estimators. We also prove the convergence of these moments, the limit matrices being functions of the $p \times p$ matrices $\Sigma_{CV,1}$ and $\Sigma_{CV,2}$, for which we prove the existence. These matrices are deterministic and depend only on the regularity parameter ϵ.

Proposition 3.5. *Assume that condition 2.1 is satisfied.*

With, for $1 \leq i \leq p$,

$$M_\theta^i = R_\theta^{-1} diag\left(R_\theta^{-1}\right)^{-2} \left\{ diag\left(R_\theta^{-1} \frac{\partial R_\theta}{\partial \theta_i} R_\theta^{-1}\right) diag\left(R_\theta^{-1}\right)^{-1} - R_\theta^{-1} \frac{\partial R_\theta}{\partial \theta_i} \right\} R_\theta^{-1},$$

we have, for all $1 \leq i, j \leq p$,

$$\frac{\partial}{\partial \theta_i} CV_\theta = \frac{1}{n} 2 y^t M_\theta^i y,$$

and

$$cov\left(\sqrt{n}\frac{\partial}{\partial \theta_i} CV_{\theta_0}, \sqrt{n}\frac{\partial}{\partial \theta_j} CV_{\theta_0} | X\right) = 2\frac{1}{n}\mathrm{Tr}\left[\left\{\mathrm{M}_{\theta_0}^{\mathrm{i}} + \left(\mathrm{M}_{\theta_0}^{\mathrm{i}}\right)^{\mathrm{t}}\right\} \mathrm{R}_{\theta_0} \left\{\mathrm{M}_{\theta_0}^{\mathrm{j}} + \left(\mathrm{M}_{\theta_0}^{\mathrm{j}}\right)^{\mathrm{t}}\right\} \mathrm{R}_{\theta_0}\right]. \quad (4)$$

Furthermore, the random trace in (4) *converges a.s. to the element* $(\Sigma_{CV,1})_{i,j}$ *of a* $p \times p$ *deterministic matrix* $\Sigma_{CV,1}$ *as* $n \to +\infty$.

We also have

$$\begin{aligned}
\mathbb{E}\left(\frac{\partial^2}{\partial \theta_i \partial \theta_j} CV_{\theta_0} | X\right) \quad = \quad & -8\frac{1}{n}\mathrm{Tr}\left\{\mathrm{diag}\left(\mathrm{R}_{\theta_0}^{-1}\right)^{-3} \mathrm{diag}\left(\mathrm{R}_{\theta_0}^{-1}\frac{\partial \mathrm{R}_{\theta_0}}{\partial \theta_{\mathrm{i}}}\mathrm{R}_{\theta_0}^{-1}\right) \mathrm{R}_{\theta_0}^{-1}\frac{\partial \mathrm{R}_{\theta_0}}{\partial \theta_{\mathrm{j}}}\mathrm{R}_{\theta_0}^{-1}\right\} \\
& +2\frac{1}{n}\mathrm{Tr}\left\{\mathrm{diag}\left(\mathrm{R}_{\theta_0}^{-1}\right)^{-2} \mathrm{R}_{\theta_0}^{-1}\frac{\partial \mathrm{R}_{\theta_0}}{\partial \theta_{\mathrm{i}}}\mathrm{R}_{\theta_0}^{-1}\frac{\partial \mathrm{R}_{\theta_0}}{\partial \theta_{\mathrm{j}}}\mathrm{R}_{\theta_0}^{-1}\right\} \\
& +6\frac{1}{n}\mathrm{Tr}\left\{\mathrm{diag}\left(\mathrm{R}_{\theta_0}^{-1}\right)^{-4} \mathrm{diag}\left(\mathrm{R}_{\theta_0}^{-1}\frac{\partial \mathrm{R}_{\theta_0}}{\partial \theta_{\mathrm{i}}}\mathrm{R}_{\theta_0}^{-1}\right) \mathrm{diag}\left(\mathrm{R}_{\theta_0}^{-1}\frac{\partial \mathrm{R}_{\theta_0}}{\partial \theta_{\mathrm{j}}}\mathrm{R}_{\theta_0}^{-1}\right) \mathrm{R}_{\theta_0}^{-1}\right\}.
\end{aligned} \quad (5)$$

Furthermore, the random trace in (5) *converges a.s. to the element* $(\Sigma_{CV,2})_{i,j}$ *of a* $p \times p$ *deterministic matrix* $\Sigma_{CV,2}$ *as* $n \to +\infty$.

In proposition 3.6, we address the asymptotic normality of CV. The conditions are, as for the consistency, identifiability and that the set of covariance functions contains only correlation functions. The convergence rate is also $\sqrt{n}$, and we have the expression of the deterministic asymptotic covariance matrix of $\sqrt{n}\hat{\theta}_{CV}$, depending only of the matrices $\Sigma_{CV,1}$ and $\Sigma_{CV,2}$ of proposition 3.5. In proposition 3.7, we prove that the asymptotic matrix $\Sigma_{CV,2}$ is positive. The minimal assumption is, as for the ML case, that the different derivative functions with respect to θ at θ_0 of the covariance function are non redundant on the set of inter-point distances covered by the random spatial sampling.

Proposition 3.6. *Assume that condition 2.1 is satisfied.*

If $\hat{\theta}_{CV}$ is consistent and if $\Sigma_{CV,2}$ is positive, then

$$\sqrt{n}\left(\hat{\theta}_{CV}-\theta_0\right)\to\mathcal{N}\left(0,\Sigma_{CV,2}^{-1}\Sigma_{CV,1}\Sigma_{CV,2}^{-1}\right)\quad as\ n\to+\infty.$$

Proposition 3.7. *Assume that condition 2.1 is satisfied and that for all $\theta\in\Theta$, $K_\theta(0)=1$.*

For $\epsilon=0$, if there does not exist $v_\lambda=(\lambda_1,...,\lambda_p)\in\mathbb{R}^p$, v_λ different from zero, so that $\sum_{k=1}^p\lambda_k\frac{\partial}{\partial\theta_k}K_{\theta_0}(v)=0$ for all $v\in\mathbb{Z}^d$, then $\Sigma_{CV,2}$ is positive.

For $\epsilon\neq 0$, we denote $D_\epsilon=\cup_{v\in\mathbb{Z}^d\setminus 0}(v+\epsilon C_{S_X})$, with $C_{S_X}=\{t_1-t_2,t_1\in S_X,t_2\in S_X\}$. If there does not exist $v_\lambda=(\lambda_1,...,\lambda_p)\in\mathbb{R}^p$, v_λ different from zero, so that $t\to\sum_{k=1}^p\lambda_k\frac{\partial}{\partial\theta_k}K_{\theta_0}(t)$ is almost surely zero on D_ϵ, with respect to the Lebesgue measure on D_ϵ, then $\Sigma_{CV,2}$ is positive.

The conclusion for ML and CV is that, for all the most classical parametric families of covariance functions, consistency and asymptotic normality hold, with deterministic positive asymptotic covariance matrices depending only on the regularity parameter ϵ. Therefore, these covariance matrices are analyzed in section 4, to address the influence of the irregularity of the spatial sampling on the ML and CV estimation.

4. Analysis of the asymptotic variances

The limit distributions of the ML and CV estimators only depend on the regularity parameter ϵ through the asymptotic covariance matrices in propositions 3.2 and 3.6. The aim of this section is to numerically study the influence of ϵ on these asymptotic covariance matrices.

The asymptotic covariance matrices of propositions 3.2 and 3.6 are expressed as functions of a.s. limits of traces of sums, products and inverses of random matrices. In the case $\epsilon=0$, for $d=1$, these matrices are deterministic Toeplitz matrices, so that the limits can be expressed using Fourier transform techniques (see [8]). In section Appendix D, we give the closed form expression of Σ_{ML}, $\Sigma_{CV,1}$ and $\Sigma_{CV,2}$ for $\epsilon=0$ and $d=1$. In the case $\epsilon\neq 0$, there does not exist, to the best of our knowledge, any random matrix technique that would give a closed form expression of Σ_{ML}, $\Sigma_{CV,1}$ and $\Sigma_{CV,2}$. Therefore, for the numerical study with $\epsilon\neq 0$, these matrices will be approximated by the random traces for large n.

4.1. The derivatives of Σ_{ML}, $\Sigma_{CV,1}$ and $\Sigma_{CV,2}$

In proposition 4.2 we show that, under the mild condition 4.1, the asymptotic covariance matrices obtained from Σ_{ML}, $\Sigma_{CV,1}$ and $\Sigma_{CV,2}$, of propositions 3.2 and 3.6, are twice differentiable with respect to ϵ. This result is useful for the numerical study of the next subsections.

Condition 4.1.

- *Condition 2.1 is satisfied.*
- *$K_\theta(t)$ and $\frac{\partial}{\partial\theta_i}K_\theta(t)$, for $1\leq i\leq p$, are three times differentiable in t for $t\neq 0$.*
- *For all $T>0$, $\theta\in\Theta$, $1\leq i\leq p$, $k\in\{1,2,3\}$, $t_1,...,t_k\in\{1,...,d\}^k$, there exists $C_T<+\infty$ so that for $|t|\geq T$,*

$$\begin{aligned}\frac{\partial}{\partial t_1},...,\frac{\partial}{\partial t_k}K_\theta(t) &\leq \frac{C_T}{1+|t|^{d+1}},\\ \frac{\partial}{\partial t_1},...,\frac{\partial}{\partial t_k}\frac{\partial}{\partial\theta_i}K_\theta(t) &\leq \frac{C_T}{1+|t|^{d+1}}.\end{aligned}\tag{6}$$

Proposition 4.2. *Assume that condition 4.1 is satisfied.*

Let us fix $1\leq i,j\leq p$. The elements $(\Sigma_{ML})_{i,j}$, $(\Sigma_{CV,1})_{i,j}$ and $(\Sigma_{CV,2})_{i,j}$ (as defined in propositions 3.2 and 3.5) are C^2 in ϵ on $[0,\frac{1}{2})$. Furthermore, with $\frac{1}{n}\mathbb{E}\{\mathrm{Tr}(\mathrm{M_{ML}})\}\to(\Sigma_{ML})_{i,j}$,

$\frac{1}{n}\mathbb{E}\left\{\mathrm{Tr}\left(\mathrm{M}_{\mathrm{CV},1}\right)\right\} \to \left(\Sigma_{CV,1}\right)_{i,j}$ *and* $\frac{1}{n}\mathbb{E}\left\{\mathrm{Tr}\left(\mathrm{M}_{\mathrm{CV},2}\right)\right\} \to \left(\Sigma_{CV,2}\right)_{i,j}$ *(propositions 3.2 and 3.5), we have, for* $\left(\Sigma\right)_{i,j}$ *being* $\left(\Sigma_{ML}\right)_{i,j}$, $\left(\Sigma_{CV,1}\right)_{i,j}$ *or* $\left(\Sigma_{CV,2}\right)_{i,j}$ *and* M *being* M_{ML}, $M_{CV,1}$ *or* $M_{CV,2}$

$$\frac{\partial^2}{\partial\epsilon^2}\left(\Sigma\right)_{i,j} = \lim_{n\to+\infty}\frac{1}{n}\mathbb{E}\left\{\frac{\partial^2}{\partial\epsilon^2}\mathrm{Tr}\left(\mathrm{M}\right)\right\}.$$

Proposition 4.2 shows that we can compute numerically the derivatives of Σ_{ML}, $\Sigma_{CV,j}$, $j = 1, 2$, with respect to ϵ by computing the derivatives of M_{ML}, $M_{CV,j}$, $j = 1, 2$, for n large. The fact that it is possible to exchange the limit in n and the derivative in ϵ was not *a priori* obvious.

In the rest of the section, we address specifically the case where $p = 1$, $d = 1$, and the law of the X_i, $1 \leq i \leq n$, is uniform on $[-1, 1]$. We focus on the case of the Matérn covariance function. In dimension one, this covariance model is parameterized by the correlation length ℓ and the smoothness parameter ν. The covariance function $K_{\ell,\nu}$ is Matérn (ℓ, ν) where

$$K_{\ell,\nu}\left(h\right) = \frac{1}{\Gamma\left(\nu\right)2^{\nu-1}}\left(2\sqrt{\nu}\frac{|h|}{\ell}\right)^{\nu} K_\nu\left(2\sqrt{\nu}\frac{|h|}{\ell}\right), \tag{7}$$

with Γ the Gamma function and K_ν the modified Bessel function of second order. See e.g [24, p.31] for a presentation of the Matérn correlation function.

4.2. Small random perturbations of the regular grid

In our study, the two true hyper-parameters (ℓ_0, ν_0) vary over $0.3 \leq \ell_0 \leq 3$ and $0.5 \leq \nu_0 \leq 5$. We will successively address the two cases where ℓ is estimated and ν is known, and where ν is estimated and ℓ is known. It is shown in section 4.1 that for both ML and CV, the asymptotic variances are regular functions of ϵ. Of course, they are even functions of ϵ, so that the quantity of interest is the ratio of the second derivative with respect to ϵ at $\epsilon = 0$ of the asymptotic variance over its value at $\epsilon = 0$. When this quantity is negative, this means that the asymptotic variance of the hyper-parameter estimator decays with ϵ, and therefore that an irregular sampling is more favorable for hyper-parameter estimation than a regular one. The second derivative is calculated exactly for ML, using the results of section Appendix D, and is approximated by finite differences for n large for CV. Proposition 4.2 ensures that this approximation is numerically consistent (because the limits in n and the derivatives in ϵ are exchangeable).

On figure 2, we show the numerical results for the estimation of ℓ. First we see that the relative improvement of the estimation due to irregularity is maximum when the true correlation length ℓ_0 is small. Indeed, the inter-observation distance being 1, a correlation length of approximatively 0.3 means that the observations are almost independent, making the estimation of the covariance very hard. Hence, the irregularity of the grid creates pairs of observations that are less independent and makes the estimation possible. For large ℓ_0, this phenomenon does not take place anymore, and thus the relative effect of the irregularity is smaller. Second, we observe that for ML the irregularity is always an advantage for estimation. This is not the case for CV, where the asymptotic variance can increase with ϵ. Finally, we can see that the two particular points $(\ell_0 = 0.5, \nu_0 = 5)$ and $(\ell_0 = 2.7, \nu_0 = 1)$ are particularly interesting and representative, since $(\ell_0 = 0.5, \nu_0 = 5)$ corresponds to hyper-parameters for which the irregularity of the sampling has a strong and favorable impact on the estimation for ML and CV, while $(\ell_0 = 2.7, \nu_0 = 1)$ corresponds to hyper-parameters for which the irregularity of the sampling has an unfavorable impact on the estimation for CV. We retain these two points for further investigation for $0 \leq \epsilon \leq 0.45$ in subsection 4.3.

On figure 3, we show the numerical results for the estimation of ν. We observe that for ℓ_0 relatively small, the asymptotic variance is an increasing function of ϵ (for small ϵ). This happens approximatively in the band $0.4 \leq \ell_0 \leq 0.6$, and for both ML and CV. This fact is not easy to interpret but it definitely gives a negative answer to the claim that irregular sampling is always better for hyper-parameter estimation than regular sampling. For $0.6 \leq \ell_0 \leq 0.8$ and $\nu_0 \geq 2$,

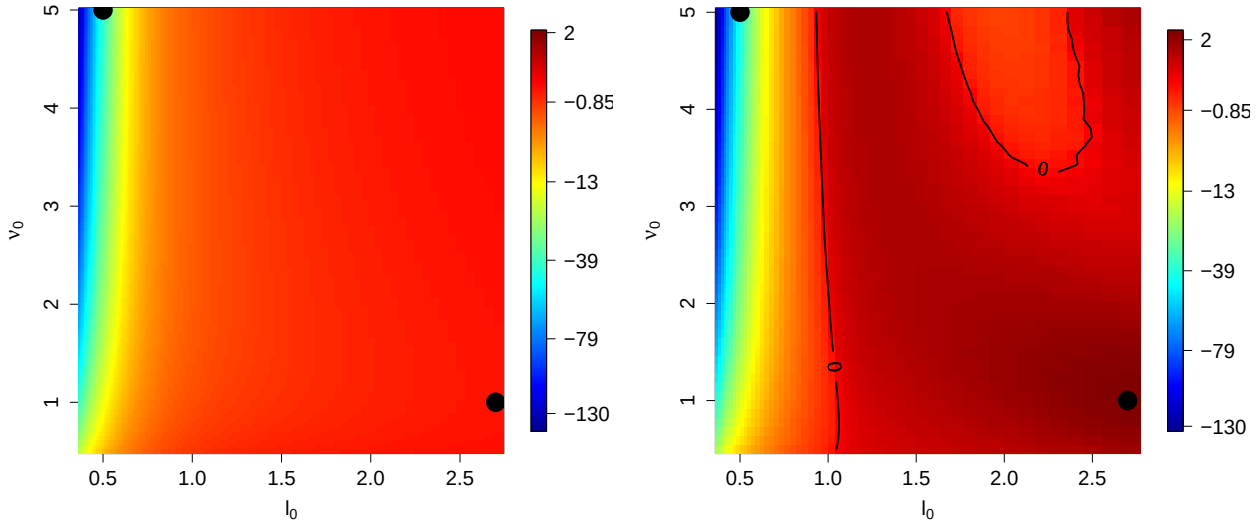

Figure 2: Local influence of ϵ for the estimation of the correlation length ℓ. Plot of the ratio of the second derivative of the asymptotic variance over its value at $\epsilon = 0$, for ML (left) and CV (right). The true covariance function is Matérn with varying ℓ_0 and ν_0. The advantage of perturbing the regular grid is maximum when the correlation length ℓ_0 small, i.e. when the observations are almost independent. The asymptotic variance always locally decreases with ϵ for ML (i.e. the second derivative at $\epsilon = 0$ is always negative) but not for CV. We retain the two particular points $(\ell_0 = 0.5, \nu_0 = 5)$ and $(\ell_0 = 2.7, \nu_0 = 1)$ for further investigation in subsection 4.3 (these are the black dots).

the relative improvement is maximum. This improvement remains significant, though smaller, for larger ℓ_0. Finally, we see the three particular points $(\ell_0 = 0.5, \nu_0 = 2.5)$, $(\ell_0 = 0.7, \nu_0 = 2.5)$ and $(\ell_0 = 2.7, \nu_0 = 2.5)$ as representative of the discussion above, and we retain them for further investigation for $0 \leq \epsilon \leq 0.45$ in subsection 4.3.

4.3. Large random perturbations of the regular grid

In this subsection, we plot the asymptotic variances of propositions 3.2 and 3.6 as functions of ϵ for $-0.45 \leq \epsilon \leq 0.45$. The asymptotic variances are even functions of ϵ. Nevertheless, they are approximated by empirical means of *iid* realizations of the random traces in propositions 3.2 and 3.5, for n large enough. Hence, the functions we plot are not exactly even. The fact that they are almost even is a graphical verification that the random fluctuations of the results of the calculations, for finite (but large) n, are very small. We also plot the second-order Taylor-series expansion given by the value at $\epsilon = 0$ and the second derivative at $\epsilon = 0$.

On figure 4, we show the numerical results for the estimation of ℓ with $(\ell_0 = 0.5, \nu_0 = 5)$. The first observation is that the asymptotic variance is slightly larger for CV than for ML. This is expected: indeed we address a well-specified case, so that the asymptotic variance of ML is the almost sure limit of the Cramer-Rao bound (the true covariance function belongs to the parametric family of covariance functions, see [3]). Therefore, this observation turns out to be true in all the subsection, and we will not comment on it anymore. We see that, for both ML and CV, the improvement of the estimation given by the irregularity of the spatial sampling is true for all values of ϵ. One can indeed gain up to a factor six for the asymptotic variances. This is explained by the reason mentioned in subsection 4.2, for ℓ_0 small, increasing ϵ yields pairs of observations that become dependent, and hence give information on the covariance structure.

On figure 5, we show the numerical results for the estimation of ℓ with $(\ell_0 = 2.7, \nu_0 = 1)$. For ML, there is a slight improvement of the estimation with the irregularity of the spatial sampling. However, for CV, there is a significant degradation of the estimation. Hence the irregularity of the spatial sampling has more relative influence on CV than on ML. Finally, the advantage of

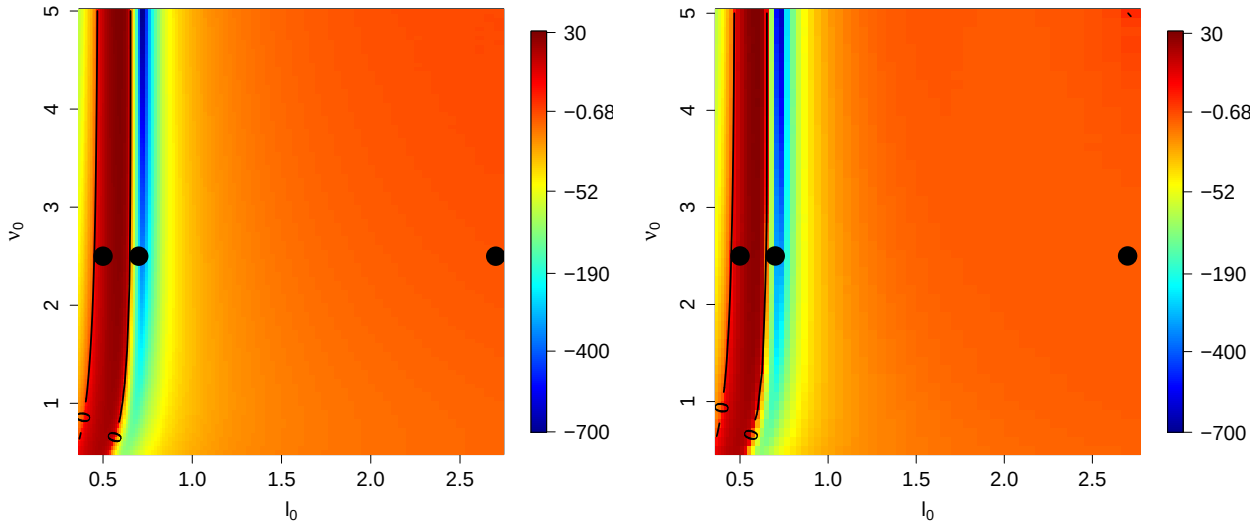

Figure 3: Same setting as figure 2, but for the estimation of ν. For approximatively $0.4 \leq \ell_0 \leq 0.6$, the estimation is damaged by locally perturbing the regular grid. For $0.6 \leq \ell_0 \leq 0.8$, the improvement of the estimation is maximum, and remains positive for larger ℓ_0. We retain the three particular points $(\ell_0 = 0.5, \nu_0 = 2.5)$, $(\ell_0 = 0.7, \nu_0 = 2.5)$ and $(\ell_0 = 2.7, \nu_0 = 2.5)$ for further investigation in subsection 4.3.

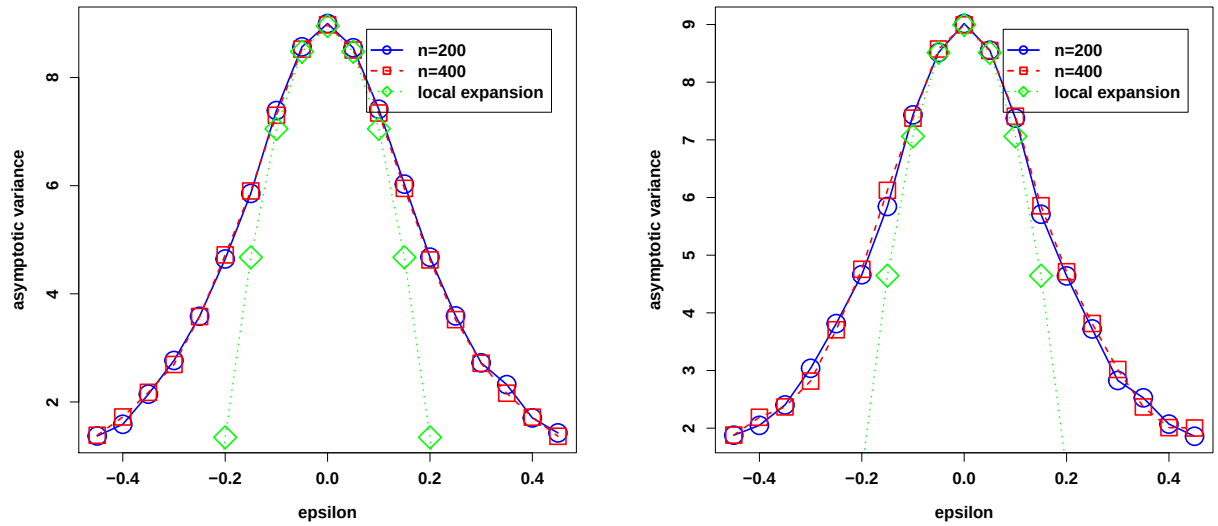

Figure 4: Global influence of ϵ for the estimation of the correlation length ℓ. Plot of the asymptotic variance for ML (left) and CV (right), calculated with varying n, and of the second order Taylor series expansion given by the value at $\epsilon = 0$ and the second derivative at $\epsilon = 0$. The true covariance function is Matérn with $\ell_0 = 0.5$ and $\nu_0 = 5$. The irregularity of the spatial sampling globally improves the estimation for both ML and CV.

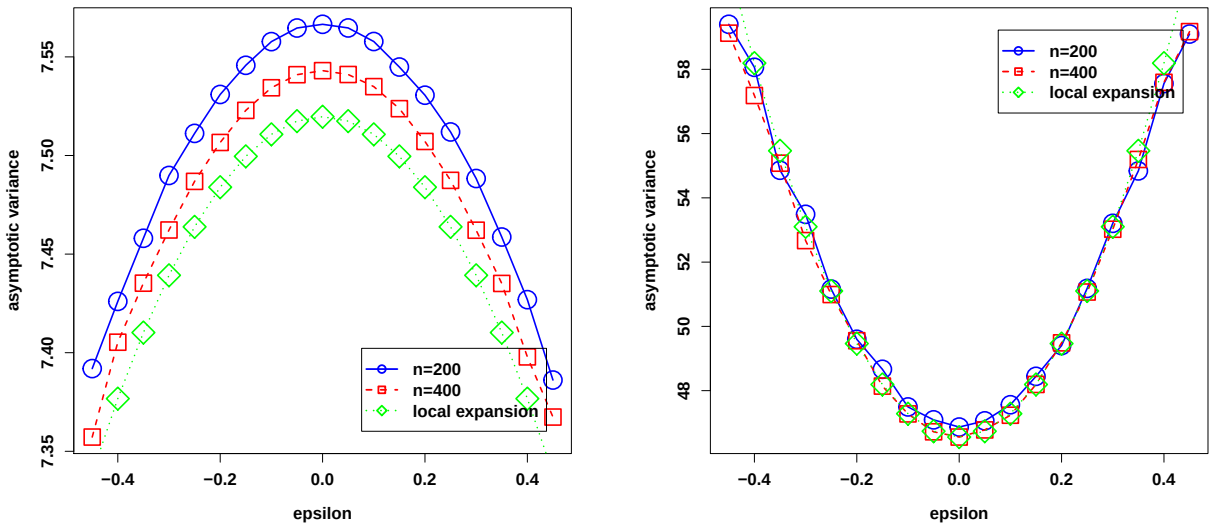

Figure 5: Same setting as in figure 4 but with $\ell_0 = 2.7$ and $\nu_0 = 1$. The irregularity of the spatial sampling slightly improves ML estimation but degrades CV estimation.

ML over CV for the estimation is by a factor seven, contrary to the case $\ell_0 = 0.5$, where this factor was close to one.

On figure 6, we show the numerical results for the estimation of ν with $(\ell_0 = 0.5, \nu_0 = 2.5)$. The numerical results are similar for ML and CV. For ϵ small, the asymptotic variance is very large, because, ℓ_0 being small, the observations are almost independent, as the observation points are further apart than the correlation length, making inference on the dependence structure very difficult. We see that, for $\epsilon = 0$, the asymptotic variance is several orders of magnitude larger than for the estimation of ℓ in figure 4, where ℓ_0 has the same value. Indeed, in the Matérn model, ν is a smoothness parameter, and its estimation is very sensitive to the absence of observation points with small spacing. Hence, naturally, for ϵ large, a threshold is reached where pairs of dependent observations start to appear, greatly reducing the asymptotic variance. However, we observe, as discussed in figure 3, that for $|\epsilon| \leq 0.2$, the asymptotic variance increases with ϵ. This non-monotony of the asymptotic variance with respect to ϵ is again a situation in which irregular sampling gives a reduced hyper-parameter estimation performance compared to regular sampling.

On figure 7, we show the numerical results for the estimation of ν with $(\ell_0 = 0.7, \nu_0 = 2.5)$. The numerical results are similar for ML and CV. Similarly to figure 6, the asymptotic variance is very large, because the observations are almost independent. However, this time the asymptotic variance is globally decreasing with ϵ. This variance is several orders of magnitude smaller for large ϵ, where pairs of dependent observations start to appear.

On figure 8, we show the numerical results for the estimation of ν with $(\ell_0 = 2.7, \nu_0 = 2.5)$. For both ML and CV, there is a global improvement of the estimation with the irregularity of the spatial sampling. Moreover, the advantage of ML over CV for the estimation, is by a factor seven, contrary to figures 6 and 7, where this factor was close to one.

4.4. Discussion

The first conclusion is that a substantial irregularity of the spatial sampling is generally an advantage for hyper-parameter estimation. Indeed, we have seen that, for ML, the asymptotic variance is always smaller for $|\epsilon| \geq 0.2$ than for $\epsilon = 0$. This is also true for CV in the case of the estimation of ν. However, for the estimation of ℓ, we can identify the cases where the

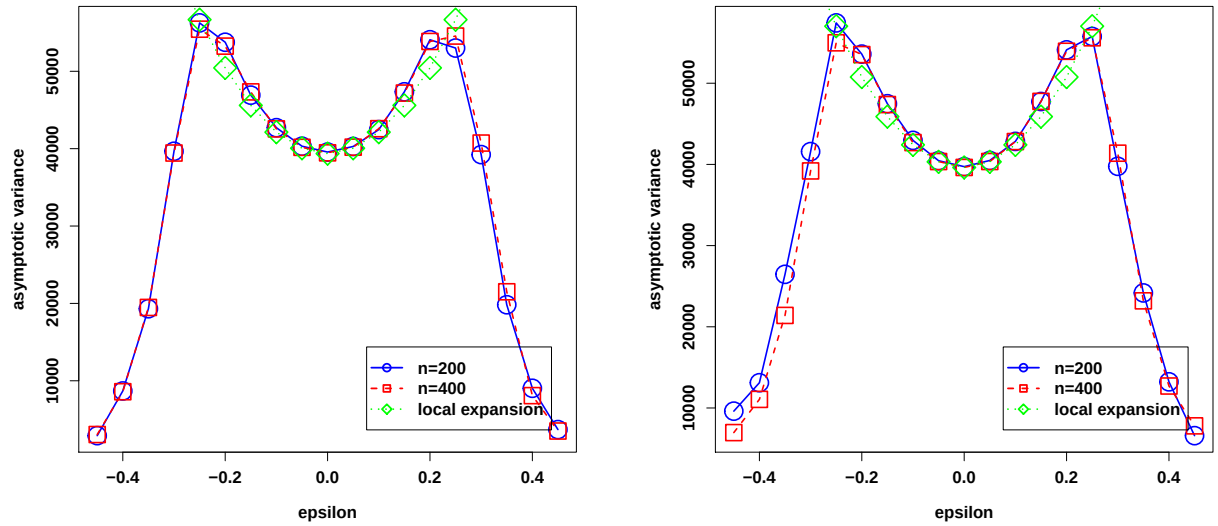

Figure 6: Same setting as in figure 4 but for the estimation of ν and with $\ell_0 = 0.5$ and $\nu_0 = 2.5$. Results are similar for ML and CV. When $\epsilon = 0$, the estimation is difficult because the observations are almost independent. The estimation is easier for ϵ large, where pairs of dependent observations start to appear. For ϵ small, the asymptotic variance increases with ϵ.

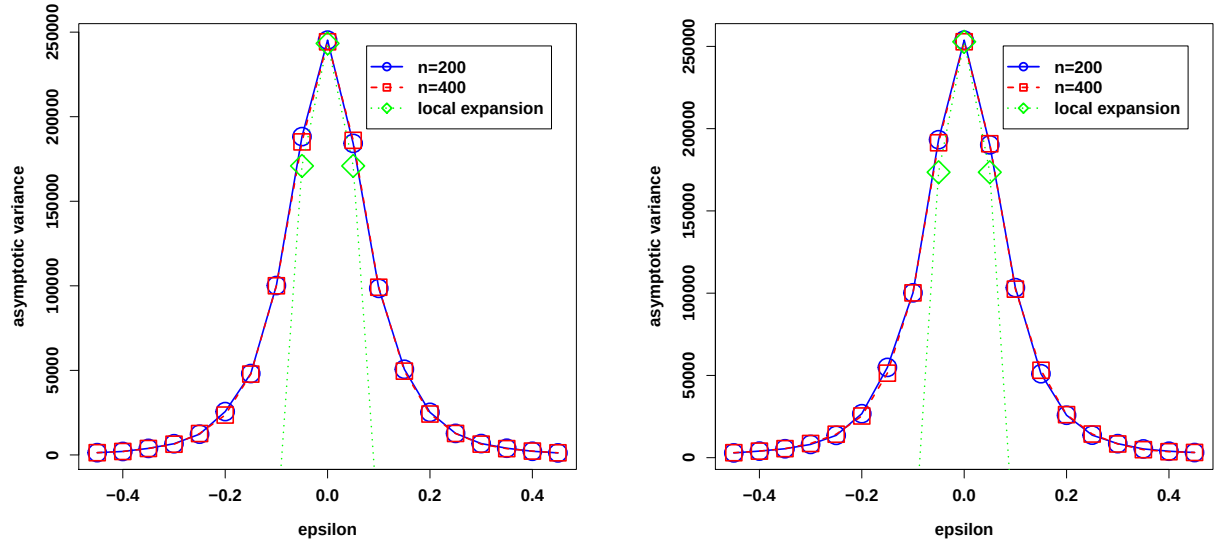

Figure 7: Same setting as in figure 4 but for the estimation of ν and with $\ell_0 = 0.7$ and $\nu_0 = 2.5$. Results are similar for ML and CV. When $\epsilon = 0$, the estimation is difficult because the observations are almost independent. The estimation is easier for ϵ large, where pairs of dependent observations start to appear.

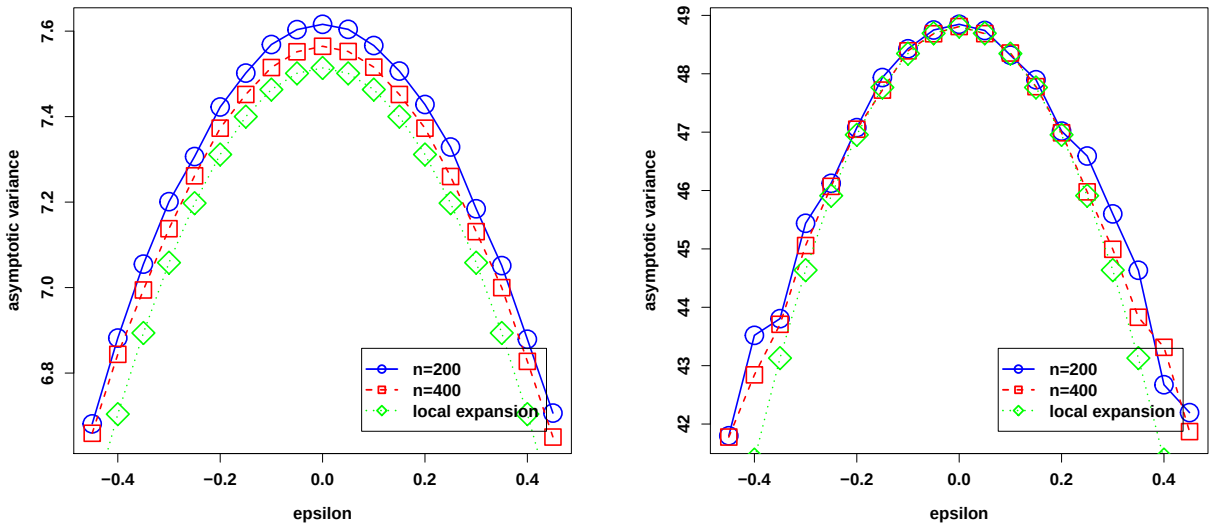

Figure 8: Same setting as in figure 4 but for the estimation of ν and with $\ell_0 = 2.7$ and $\nu_0 = 2.5$. For both ML and CV, there is a global improvement of the estimation with the irregularity of the spatial sampling. ML has a substantial advantage over CV for the estimation.

asymptotic variance is globally increasing with ϵ. This can be due to the fact that the Leave-One-Out errors in the CV functional are unnormalized. Hence, with $\epsilon \neq 0$, roughly speaking, error terms concerning observation points with close neighbors are small, while error terms concerning observation points without close neighbors are large. Hence, the CV functional mainly depends on the large error terms and hence has a larger variance.

Nevertheless, the second conclusion is that locally perturbing a regular observation grid can damage the estimation for both ML (figure 3) and CV (figures 2 and 3). This is the most important practical observation that follows for our detailed analysis of the asymptotic variances of the hyper-parameter estimators.

5. Influence of hyper-parameter misspecification on prediction

In proposition 5.1, we show that the misspecification of correlation hyper-parameters has an asymptotic influence on the prediction errors. Indeed, the difference of the asymptotic Leave-One-Out mean square errors, between incorrect and correct hyper-parameters, is lower and upper bounded by finite positive constants times the integrated square difference between the two correlation functions.

Proposition 5.1. *Assume that condition 2.1 is satisfied and that for all $\theta \in \Theta$, $K_\theta(0) = 1$.*

Let, for $1 \leq i \leq n$, $\hat{y}_{i,\theta}(y_{-i}) := \mathbb{E}_{\theta|X}(y_i|y_1, ..., y_{i-1}, y_{i+1}, ..., y_n)$ be the Kriging Leave-One-Out prediction of y_i with covariance hyper-parameters θ. We then denote

$$D_p(\theta, \theta_0) := \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\{y_i - \hat{y}_{i,\theta}(y_{-i})\}^2\right] - \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\{y_i - \hat{y}_{i,\theta_0}(y_{-i})\}^2\right].$$

Then there exists constant $0 < A < B < +\infty$ so that, for $\epsilon = 0$

$$A\sum_{v\in\mathbb{Z}^d}\{K_\theta(v) - K_{\theta_0}(v)\}^2 \leq \varliminf_{n\to+\infty} D_p(\theta, \theta_0)$$

and

$$\overline{\lim_{n\to+\infty}} D_p(\theta,\theta_0) \leq B \sum_{v\in\mathbb{Z}^d} \left\{K_\theta(v) - K_{\theta_0}(v)\right\}^2.$$

For $\epsilon \neq 0$, we denote $D_\epsilon = \cup_{v\in\mathbb{Z}^d\setminus 0}\,(v + \epsilon C_{S_X})$, with $C_{S_X} = \{t_1 - t_2, t_1 \in S_X, t_2 \in S_X\}$. Then

$$A \int_{D_\epsilon} \left\{K_\theta(t) - K_{\theta_0}(t)\right\}^2 dt \leq \underline{\lim_{n\to+\infty}} D_p(\theta,\theta_0)$$

and

$$\overline{\lim_{n\to+\infty}} D_p(\theta,\theta_0) \leq B \int_{D_\epsilon} \left\{K_\theta(t) - K_{\theta_0}(t)\right\}^2 dt$$

Proof. The minoration is proved in the proof of proposition 3.4. The majoration is obtained with similar techniques. □

6. Conclusion

We have considered an increasing-domain asymptotic framework to address the influence of the irregularity of the spatial sampling on the estimation of the covariance hyper-parameters. This framework is based on a random sequence of observation points, for which the deviation from the regular grid is controlled by a single scalar regularity parameter ϵ.

We have proved consistency and asymptotic normality for the ML and CV estimators, under rather minimal conditions. The asymptotic variances are deterministic functions of the regularity parameter only. Hence, it is the natural tool to assess the influence of the irregularity of the spatial sampling on the ML and CV estimators.

This is carried out by means of an exhaustive study of the Matérn model. We put into evidence that the irregularity of the spatial sampling is generally an advantage for estimation. However, we show that there exist cases for ML when disrupting a regular spatial sampling can on the contrary damage estimation. In the CV case, estimation can be very strongly damaged for a strong random perturbation of the grid.

Hence, the overall conclusion is that we definitely give a negative answer to the claim that irregular sampling is always better for hyper-parameter estimation than regular sampling. The influence of the regularity of the spatial sampling on the covariance hyper-parameter estimation remains a non trivial problem.

The CV criterion we have studied is a mean square error criterion. This is a classical CV criterion that is used, for instance, in [20] when the CV and ML estimations of the hyper-parameters are compared. We have shown that this CV estimator can have a considerably larger asymptotic variance than the ML estimator, on the one hand, and can be sensitive to the irregularity of the spatial sampling, on the other hand. Although this estimator performs better than ML in cases of model misspecification [3], further research may aim at studying alternative CV criteria that would have a better performance in the well-specified case.

Other CV criteria are proposed in the literature, for instance the LOO log-predictive probability in [17] and the Geisser's predictive mean square error in [25]. It would be interesting to study, in the framework of this paper, the increasing-domain asymptotics for these estimators and the influence of the irregularity of the spatial sampling.

In section 4, we pointed out that, when the spatial sampling is irregular, the mean square error CV criterion could be composed of LOO errors with heterogeneous variances, which increases the CV estimation variance. Methods to normalize the LOO errors would be an interesting research direction to explore.

Acknowledgments

I would like to thank my advisors Josselin Garnier (University Paris Diderot) and Jean-Marc Martinez (French Alternative Energies and Atomic Energy Commission - Nuclear Energy Division at CEA-Saclay, DEN, DM2S, STMF, LGLS) for their advices and suggestions.

Appendix A. Proofs for section 3

In the proofs, we distinguish three probability spaces.

$(\Omega_X, \mathcal{F}_X, P_X)$ is the probability space associated with the random perturbation of the regular grid. $(X_i)_{i\in\mathbb{N}^*}$ is a sequence of iid S_X-valued random variables defined on $(\Omega_X, \mathcal{F}_X, P_X)$, with distribution $\mathcal{L}_X$. We denote by ω_X an element of Ω_X.

$(\Omega_Y, \mathcal{F}_Y, P_Y)$ is the probability space associated with the Gaussian process. Y is a centered Gaussian process with covariance function K_{θ_0} defined on $(\Omega_Y, \mathcal{F}_Y, P_Y)$. We denote by ω_Y an element of Ω_Y.

$(\Omega, \mathcal{F}, \mathbb{P})$ is the product space $(\Omega_X \times \Omega_Y, \mathcal{F}_X \otimes \mathcal{F}_Y, P_X \times P_Y)$. We denote by ω an element of Ω.

All the random variables in the proofs can be defined relatively to the product space $(\Omega, \mathcal{F}, \mathbb{P})$. Hence, all the probabilistic statements in the proofs hold with respect to this product space, unless it is stated otherwise.

In the proofs, when $(f_n)_{n\in\mathbb{N}^*}$ is a sequence of real functions of $X = (X_i)_{i=1}^n$, f_n is also a sequence of real random variables on $(\Omega_X, \mathcal{F}_X, P_X)$. When we write that f_n is bounded uniformly in n and x, we mean that there exists a finite constant K so that $\sup_n \sup_{x\in S_X^n} |f_n(x)| \leq K$. We then have that f_n is bounded P_X-a.s., i.e $\sup_n f_n \leq K$ for a.e. $\omega_X \in \Omega_X$. We may also write that f_n is lower-bounded uniformly in n and x when there exist $a > 0$ so that $\inf_n \inf_{x\in S_X^n} f_n(x) \geq a$. When f_n also depends on θ, we say that f_n is bounded uniformly in n, x and θ when $\sup_{\theta\in\Theta} f_n$ is bounded uniformly in n and x. We also say that f_n is lower-bounded uniformly in n, x and θ when $\inf_{\theta\in\Theta} f_n$ is lower-bounded uniformly in n and x.

When we write that f_n converges to zero uniformly in x, we mean that $\sup_{x\in S_X^n} |f_n(x)| \to_{n\to+\infty} 0$. One then have that f_n converges to zero P_X-a.s. When f_n also depends on θ, we say that f_n converges to zero uniformly in n, x and θ when $\sup_{\theta\in\Theta} f_n$ converges to zero uniformly in n and x.

When f_n is a sequence of real functions of X and Y, f_n is also a sequence of real random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. When we say that f_n is bounded in probability conditionally to $X = x$ and uniformly in x, we mean that, for every $\epsilon > 0$, there exist M, N so that $\sup_{n\geq N} \sup_{x\in S_X^n} \mathbb{P}(|f_n| \geq M | X = x) \leq \epsilon$. One then have that f_n is bounded in probability (defined on the product space).

Appendix A.1. Proof of proposition 3.1

Proof. We show that there exist sequences of random variables, defined on $(\Omega_X, \mathcal{F}_X, P_X)$, D_{θ,θ_0} and D_{2,θ,θ_0} (functions of n and X), so that $\sup_\theta |(L_\theta - L_{\theta_0}) - D_{\theta,\theta_0}| \to_p 0$ (in probability of the product space) and $D_{\theta,\theta_0} \geq B D_{2,\theta,\theta_0}$ P_X-a.s. for a constant $B > 0$. We then show that there exists $D_{\infty,\theta,\theta_0}$, a deterministic function of θ, θ_0 only, so that $\sup_\theta |D_{2,\theta,\theta_0} - D_{\infty,\theta,\theta_0}| = o_p(1)$ and for any $\alpha > 0$,

$$\inf_{|\theta-\theta_0|\geq\alpha} D_{\infty,\theta,\theta_0} > 0. \tag{A.1}$$

This implies consistency.

We have $L_\theta = \frac{1}{n}\log\{\det(R_\theta)\} + \frac{1}{n} y^t R_\theta^{-1} y$. The eigenvalues of R_θ and R_θ^{-1} being bounded uniformly in n and x (lemma Appendix C.5), $var(L_\theta | X = x)$ converges to 0 uniformly in x, and so $L_\theta - \mathbb{E}(L_\theta | X)$ converges in probability $\mathbb{P}$ to zero.

Then, with $z = R_{\theta_0}^{-\frac{1}{2}} y$,

$$\begin{aligned}
\sup_{k\in\{1,\dots,p\},\theta\in\Theta} \left|\frac{\partial L_\theta}{\partial\theta_k}\right| &= \sup_{k\in\{1,\dots,p\},\theta\in\Theta} \frac{1}{n}\left\{\mathrm{Tr}\left(\mathrm{R}_\theta^{-1}\frac{\partial \mathrm{R}_\theta}{\partial\theta_\mathrm{k}}\right) + \mathrm{z}^\mathrm{t}\mathrm{R}_{\theta_0}^{\frac{1}{2}}\mathrm{R}_\theta^{-1}\frac{\partial \mathrm{R}_\theta}{\partial\theta_\mathrm{k}}\mathrm{R}_\theta^{-1}\mathrm{R}_{\theta_0}^{\frac{1}{2}}\mathrm{z}\right\} \\
&\leq \sup_{k\in\{1,\dots,p\},\theta} \left\{ max\left(||R_\theta^{-1}|| \left|\left|\frac{\partial R_\theta}{\partial\theta_k}\right|\right|, ||R_{\theta_0}|| \, ||R_{\theta^{-2}}|| \left|\left|\frac{\partial R_\theta}{\partial\theta_k}\right|\right| \right) \right\} \left(1 + \frac{1}{n}|z|^2\right),
\end{aligned}$$

and is hence bounded in probability conditionally to $X = x$, uniformly in x, because of lemma Appendix C.5 and the fact that $z \sim \mathcal{N}(0, I_n)$ given $X = x$.

Because of the simple convergence and the boundedness of the derivatives, $\sup_\theta |L_\theta - \mathbb{E}(L_\theta|X)| \to_p 0$. We then denote $D_{\theta,\theta_0} := \mathbb{E}(L_\theta|X) - \mathbb{E}(L_{\theta_0}|X)$. We then have $\sup_\theta |(L_\theta - L_{\theta_0}) - D_{\theta,\theta_0}| \to_p 0$.

We have $\mathbb{E}(L_\theta|X) = \frac{1}{n}\log\{\det(R_\theta)\} + \frac{1}{n}\mathrm{Tr}\left(\mathrm{R}_\theta^{-1}\mathrm{R}_{\theta_0}\right)$ and hence, P_X-a.s.

$$\begin{aligned} D_{\theta,\theta_0} &= \frac{1}{n}\log\{\det(R_\theta)\} + \frac{1}{n}\mathrm{Tr}\left(\mathrm{R}_\theta^{-1}\mathrm{R}_{\theta_0}\right) - \frac{1}{\mathrm{n}}\log\{\det(\mathrm{R}_{\theta_0})\} - 1 \\ &= \frac{1}{n}\sum_{i=1}^{n}\left[-\log\left\{\phi_i\left(R_{\theta_0}^{\frac{1}{2}}R_\theta^{-1}R_{\theta_0}^{\frac{1}{2}}\right)\right\} + \phi_i\left(R_{\theta_0}^{\frac{1}{2}}R_\theta^{-1}R_{\theta_0}^{\frac{1}{2}}\right) - 1\right]. \end{aligned}$$

Using proposition Appendix C.4 and lemma Appendix C.5, there exists $0 < a < b < +\infty$ so that for all x, n, θ, $a < \phi_i\left(R_{\theta_0}^{\frac{1}{2}}R_\theta^{-1}R_{\theta_0}^{\frac{1}{2}}\right) < b$. We denote $f(t) = -\log(t) + t - 1$. As f is minimal in 1, $f'(1) = 0$ and $f''(1) = 1$, there exists $A > 0$ so that, for $t \in [a,b]$, $f(t)$ is larger than $A(t-1)^2$. Then,

$$\begin{aligned} D_{\theta,\theta_0} &\geq A\frac{1}{n}\sum_{i=1}^{n}\left\{1 - \phi_i\left(R_{\theta_0}^{\frac{1}{2}}R_\theta^{-1}R_{\theta_0}^{\frac{1}{2}}\right)\right\}^2 \\ &= A\frac{1}{n}\mathrm{Tr}\left\{\left(\mathrm{I} - \mathrm{R}_{\theta_0}^{\frac{1}{2}}\mathrm{R}_\theta^{-1}\mathrm{R}_{\theta_0}^{\frac{1}{2}}\right)^2\right\} \\ &= A\frac{1}{n}\left|R_\theta^{-\frac{1}{2}}\left(R_\theta - R_{\theta_0}\right)R_\theta^{-\frac{1}{2}}\right|^2. \end{aligned}$$

Then, as the eigenvalues of $R_\theta^{-\frac{1}{2}}$ are larger than $c > 0$, uniformly in n, x and θ, and with $|MN|^2 \geq \inf_i \phi_i^2(M)|N|^2$ for M symmetric positive, we obtain, for some $B > 0$, and uniformly in n, x and θ,

$$D_{\theta,\theta_0} \geq B|R_\theta - R_{\theta_0}|^2 := BD_{2,\theta,\theta_0}.$$

For $\epsilon = 0$, D_{2,θ,θ_0} is deterministic and converges to

$$D_{\infty,\theta,\theta_0} := \sum_{v\in\mathbb{Z}^d}\{K_\theta(v) - K_{\theta_0}(v)\}^2. \tag{A.2}$$

$D_{\infty,\theta,\theta_0}$ est continuous in θ because the series of term $\sup_\theta |K_\theta(v)|^2$, $v \in \mathbb{Z}^d$ is summable using (1) and lemma Appendix C.1. Hence, if there exists $\alpha > 0$, $\inf_{|\theta-\theta_0|\geq\alpha} D_{\infty,\theta,\theta_0} = 0$, we can, using a compacity and continuity argument, have $\theta_\infty \neq \theta_0$ so that (A.2) is null. Hence we showed (A.1) by contradiction, which shows the proposition for $\epsilon = 0$.

For $\epsilon \neq 0$, $D_{2,\theta,\theta_0} = \frac{1}{n}\mathrm{Tr}\left\{\left(\mathrm{R}_\theta - \mathrm{R}_{\theta_0}\right)^2\right\}$. With fixed θ, using proposition Appendix C.7, D_{2,θ,θ_0} converges in P_X-probability to $D_{\infty,\theta,\theta_0} := \lim_{n\to\infty} E_X(D_{2,\theta,\theta_0})$. The eigenvalues of the $\frac{\partial R_\theta}{\partial\theta_i}$, $1 \leq i \leq n$, being bounded uniformly in n, θ, x, the partial derivatives with respect to θ of D_{2,θ,θ_0} are uniformly bounded in n, θ and x. Hence $\sup_\theta |D_{2,\theta,\theta_0} - D_{\infty,\theta,\theta_0}| = o_p(1)$. Then

$$D_{\infty,\theta,\theta_0} = \lim_{n\to+\infty}\frac{1}{n}\sum_{1\leq i,j\leq n, i\neq j}\left[\int_{\epsilon C_{S_X}}\{K_\theta(v_i - v_j + t) - K_{\theta_0}(v_i - v_j + t)\}^2 f_T(t)\,dt\right] + \{K_\theta(0) - K_{\theta_0}(0)\}^2,$$

with $f_T(t)$ the probability density function of $\epsilon(X_i - X_j)$, $i \neq j$. We then show,

$$\begin{aligned} D_{\infty,\theta,\theta_0} &= \sum_{v\in\mathbb{Z}^d\setminus 0}\left[\int_{\epsilon C_{S_X}}\{K_\theta(v+t) - K_{\theta_0}(v+t)\}^2 f_T(t)\,dt\right] + \{K_\theta(0) - K_{\theta_0}(0)\}^2 \\ &= \int_{D_\epsilon}\{K_\theta(t) - K_{\theta_0}(t)\}^2 f_T(t)\,dt + \{K_\theta(0) - K_{\theta_0}(0)\}^2. \end{aligned} \tag{A.3}$$

As $\sup_\theta |K_\theta(t)|^2$ is summable on D_ϵ, using (1), $D_{\infty,\theta,\theta_0}$ is continuous. Hence, if there exists $\alpha > 0$ so that $\inf_{|\theta-\theta_0|\geq\alpha} D_{\infty,\theta,\theta_0} = 0$, we can, using a compacity and continuity argument, show that there exists $\theta_\infty \neq \theta_0$ so that (A.3) is null. Hence we proved (A.1) by contradiction which proves the proposition for $\epsilon \neq 0$.

□

Appendix A.2. proof of Proposition 3.2

Proof. For $1 \leq i,j \leq p$, we use proposition Appendix C.7 to show that $\frac{1}{n}\mathrm{Tr}\left(\mathrm{R}^{-1}\frac{\partial \mathrm{R}}{\partial \theta_{\mathrm{i}}}\mathrm{R}^{-1}\frac{\partial \mathrm{R}}{\partial \theta_{\mathrm{j}}}\right)$ has a P_X-almost sure limit as $n \to +\infty$.

We calculate $\frac{\partial}{\partial \theta_i}L_\theta = \frac{1}{n}\left\{\mathrm{Tr}\left(\mathrm{R}_\theta^{-1}\frac{\partial \mathrm{R}_\theta}{\partial \theta_{\mathrm{i}}}\right) - \mathrm{y}^{\mathrm{t}}\mathrm{R}_\theta^{-1}\frac{\partial \mathrm{R}_\theta}{\partial \theta_{\mathrm{i}}}\mathrm{R}_\theta^{-1}\mathrm{y}\right\}$. We use proposition Appendix C.9 with $M_i = R_\theta^{-1}\frac{\partial R_\theta}{\partial \theta_i}$ and $N_i = -R_\theta^{-1}\frac{\partial R_\theta}{\partial \theta_i}R_\theta^{-1}$, together with proposition Appendix C.7, to show that

$$\sqrt{n}\frac{\partial}{\partial \theta}L_{\theta_0} \to \mathcal{N}\left(0, 2\Sigma_{ML}\right).$$

We calculate

$$\begin{aligned}\frac{\partial^2}{\partial \theta_i \partial \theta_j}L_{\theta_0} &= \frac{1}{n}\mathrm{Tr}\left(-\mathrm{R}^{-1}\frac{\partial \mathrm{R}}{\partial \theta_{\mathrm{i}}}\mathrm{R}^{-1}\frac{\partial \mathrm{R}}{\partial \theta_{\mathrm{j}}} + \mathrm{R}^{-1}\frac{\partial^2 \mathrm{R}}{\partial \theta_{\mathrm{i}}\partial \theta_{\mathrm{j}}}\right) \\ &+ \frac{1}{n}y^t\left(2R^{-1}\frac{\partial R}{\partial \theta_i}R^{-1}\frac{\partial R}{\partial \theta_j}R^{-1} - R^{-1}\frac{\partial^2 R}{\partial \theta_i \partial \theta_j}R^{-1}\right)y.\end{aligned}$$

Hence, using proposition Appendix C.8, $\frac{\partial^2}{\partial \theta^2}L_{\theta_0}$ converges to Σ_{ML} in the mean square sense (on the product space).

Finally, $\frac{\partial^3}{\partial \theta_i \partial \theta_j \partial \theta_k}L_{\tilde{\theta}}$ can be written as $\frac{1}{n}\left\{\mathrm{Tr}\left(\mathrm{M}_{\tilde{\theta}}\right) + \mathrm{z}^{\mathrm{t}}\mathrm{N}_{\tilde{\theta}}\mathrm{z}\right\}$, where $M_{\tilde{\theta}}$ and $N_{\tilde{\theta}}$ are sums of matrices of $\mathcal{M}_{\tilde{\theta}}$ (proposition Appendix C.7) and where z depends on X and Y and $(z|X) = \mathcal{N}(0, I_n)$. Hence, the singular values of $M_{\tilde{\theta}}$ and $N_{\tilde{\theta}}$ are bounded uniformly in $\tilde{\theta}$, n and x, and so $\sup_{i,j,k\tilde{\theta}}\frac{\partial^3}{\partial \theta_i \partial \theta_j \partial \theta_k}L_{\tilde{\theta}}$ is bounded by $a + b\frac{1}{n}|z|^2$, with constant $a, b < +\infty$ and is hence bounded in probability. Hence we apply proposition Appendix C.10 to conclude.

□

Appendix A.3. Proof of proposition 3.3

Proof. We firstly prove the proposition in the case $p = 1$, when Σ_{ML} is a scalar. We then show how to generalize the proposition to the case $p > 1$.

For $p = 1$ we have seen that $\frac{1}{n}\mathrm{Tr}\left(\mathrm{R}_{\theta_0}^{-1}\frac{\partial \mathrm{R}_{\theta_0}}{\partial \theta}\mathrm{R}_{\theta_0}^{-1}\frac{\partial \mathrm{R}_{\theta_0}}{\partial \theta}\right) \to_{\mathrm{P_X}} \Sigma_{\mathrm{ML}}$. Then

$$\begin{aligned}\frac{1}{n}\mathrm{Tr}\left(\mathrm{R}_{\theta_0}^{-1}\frac{\partial \mathrm{R}_{\theta_0}}{\partial \theta}\mathrm{R}_{\theta_0}^{-1}\frac{\partial \mathrm{R}_{\theta_0}}{\partial \theta}\right) = \frac{1}{\mathrm{n}}\mathrm{Tr}\left(\mathrm{R}_{\theta_0}^{-\frac{1}{2}}\frac{\partial \mathrm{R}_{\theta_0}}{\partial \theta}\mathrm{R}_{\theta_0}^{-\frac{1}{2}}\mathrm{R}_{\theta_0}^{-\frac{1}{2}}\frac{\partial \mathrm{R}_{\theta_0}}{\partial \theta}\mathrm{R}_{\theta_0}^{-\frac{1}{2}}\right) &= \left|R_{\theta_0}^{-\frac{1}{2}}\frac{\partial R_{\theta_0}}{\partial \theta}R_{\theta_0}^{-\frac{1}{2}}\right|^2 \\ &\geq \inf_{i,n,x}\phi_i\left(R_{\theta_0}^{-\frac{1}{2}}\right)^4\left|\frac{\partial R_{\theta_0}}{\partial \theta}\right|^2.\end{aligned}$$

By lemma Appendix C.5, there exists $a > 0$ so that $\inf_{i,n,x}\phi_i\left(R_{\theta_0}^{-\frac{1}{2}}\right)^4 \geq a$. We then show, similarly to the proof of proposition 3.1, that the limit of $\left|\frac{\partial R_{\theta_0}}{\partial \theta}\right|^2$ is positive.

We now address the case $p > 1$. Let $v_\lambda = \lambda_1, ..., \lambda_p \in \mathbb{R}^p$, v_λ different from zero. We define the model $\{K_\delta, \delta \in [\delta_{inf}, \delta_{sup}]\}$, with $\delta_{inf} < 0 < \delta_{sup}$ by $K_\delta = K_{(\theta_0)_1 + \delta\lambda_1, ..., (\theta_0)_p + \delta\lambda_p}$. Then $K_{\delta=0} = K_{\theta_0}$. We have $\frac{\partial}{\partial \delta}K_{\delta=0}(t) = \sum_{k=1}^p \lambda_k \frac{\partial}{\partial \theta_k}K_{\theta_0}(t)$, so the model $\{K_\delta, \delta \in [\delta_{inf}, \delta_{sup}]\}$ verifies the hypotheses of the proposition for $p = 1$. Hence, the $\mathbb{P}$-mean square limit of $\frac{\partial^2}{\partial \delta^2}L_{\delta=0}$ is positive. We conclude with $\frac{\partial^2}{\partial \delta^2}L_{\delta=0} = v_\lambda^t\left(\frac{\partial^2}{\partial \theta^2}L_{\theta_0}\right)v_\lambda$.

□

Appendix A.4. Proof of proposition 3.4

Proof. We will show that there exists a sequence of random variables defined on $(\Omega_X, \mathcal{F}_X, P_X)$ D_{θ,θ_0} so that $\sup_\theta |(CV_\theta - CV_{\theta_0}) - D_{\theta,\theta_0}| \to_p 0$ and $C > 0$ so that P_X-a.s.

$$D_{\theta,\theta_0} \geq C|R_\theta - R_{\theta_0}|^2. \tag{A.4}$$

The proof of the proposition is then carried out similarly to the proof of proposition 3.1.

We firstly show, similarly to the proof of proposition 3.1 that $\sup_\theta |CV_\theta - \mathbb{E}(CV_\theta|X)| \to_p 0$. We then denote $D_{\theta,\theta_0} = \mathbb{E}(CV_\theta|X) - \mathbb{E}(CV_{\theta_0}|X)$. We decompose, for all $i \in \{1,...,n\}$, with P_i the matrix that exchanges lines 1 and i of a matrix,

$$P_i R_\theta P_i^t = \begin{pmatrix} 1 & r_{i,\theta}^t \\ r_{i,\theta} & R_{-i,\theta} \end{pmatrix}.$$

The conditional laws being independent on the numbering of the observations, we have, using the Kriging equations

$$\begin{aligned} D_{\theta,\theta_0} &= \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left\{ \left(r_{i,\theta}^t R_{-i,\theta}^{-1} y_{-i} - r_{i,\theta_0}^t R_{-i,\theta_0}^{-1} y_{-i} \right)^2 |X \right\} \\ &= \frac{1}{n}\sum_{i=1}^n \left(r_{i,\theta}^t R_{-i,\theta}^{-1} - r_{i,\theta_0}^t R_{-i,\theta_0}^{-1} \right) R_{-i,\theta_0} \left(R_{-i,\theta}^{-1} r_{i,\theta} - R_{-i,\theta_0}^{-1} r_{i,\theta_0} \right). \end{aligned}$$

Similarly to lemma Appendix C.5, it can be shown that the eigenvalues of R_{-i,θ_0} are larger than a constant $A > 0$, uniformly in n and x. Then

$$D_{\theta,\theta_0} \geq A\frac{1}{n}\sum_{i=1}^n \left|\left| \left(r_{i,\theta}^t R_{-i,\theta}^{-1} - r_{i,\theta_0}^t R_{-i,\theta_0}^{-1} \right) \right|\right|^2.$$

Using the virtual Cross Validation equations [18, ch.5.2], the vector $R_{-i,\theta}^{-1} r_{i,\theta}$ is the vector of the $\frac{\left(R_\theta^{-1}\right)_{i,j}}{\left(R_\theta^{-1}\right)_{i,i}}$ for $1 \leq j \leq n$, $j \neq i$. Hence P_X-a.s.

$$\begin{aligned} D_{\theta,\theta_0} &\geq A\frac{1}{n}\sum_{i=1}^n \sum_{j\neq i} \left\{ \frac{\left(R_\theta^{-1}\right)_{i,j}}{\left(R_\theta^{-1}\right)_{i,i}} - \frac{\left(R_{\theta_0}^{-1}\right)_{i,j}}{\left(R_{\theta_0}^{-1}\right)_{i,i}} \right\}^2 \\ &= A\frac{1}{n}\left| diag\left(R_\theta^{-1}\right)^{-1} R_\theta^{-1} - diag\left(R_{\theta_0}^{-1}\right)^{-1} R_{\theta_0}^{-1} \right|^2 \\ &\geq AB\frac{1}{n}\left| diag\left(R_{\theta_0}^{-1}\right) diag\left(R_\theta^{-1}\right)^{-1} R_\theta^{-1} - R_{\theta_0}^{-1} \right|^2, \quad \text{with } B = \inf_{i,n,x} \phi_i^2 \left\{ diag\left(R_{\theta_0}^{-1}\right)^{-1} \right\},\ B > 0. \end{aligned}$$

The eigenvalues of $diag\left(R_{\theta_0}^{-1}\right) diag\left(R_\theta^{-1}\right)^{-1}$ are bounded between $a > 0$ and $b < \infty$ uniformly in n and x. Hence we have, with D_λ, the diagonal matrix with values $\lambda_1, ..., \lambda_n$,

$$\begin{aligned} D_{\theta,\theta_0} &\geq AB \inf_{a\leq\lambda_1,...,\lambda_n\leq b} \left| D_\lambda R_\theta^{-1} - R_{\theta_0}^{-1} \right|^2 \\ &\geq ABC \inf_{a\leq\lambda_1,...,\lambda_n\leq b} |D_{\frac{1}{\lambda}} R_\theta - R_{\theta_0}|^2, \quad \text{using [8] theorem 2.1} \\ &\geq ABC \inf_{\lambda_1,...,\lambda_n} |D_\lambda R_\theta - R_{\theta_0}|^2, \end{aligned}$$

with $C = \frac{1}{b}\inf_{n,x,\theta} \frac{1}{||R_\theta||^2} \frac{1}{||R_{\theta_0}||^2}$, $C > 0$. Then

$$\begin{aligned} D_{\theta,\theta_0} &\geq ABC\frac{1}{n} \inf_{\lambda_1,...,\lambda_n} \sum_{i,j=1}^n \left(\lambda_i R_{\theta,i,j} - R_{\theta_0,i,j} \right)^2 \\ &= ABC\frac{1}{n}\sum_{i=1}^n \inf_\lambda \sum_{j=1}^n \left(\lambda R_{\theta,i,j} - R_{\theta_0,i,j} \right)^2 \\ &= ABC\frac{1}{n}\sum_{i=1}^n \inf_\lambda \left\{ (\lambda - 1)^2 + \sum_{j\neq i} \left(\lambda R_{\theta,i,j} - R_{\theta_0,i,j} \right)^2 \right\}. \end{aligned}$$

Lemma Appendix A.1. *For any $a_1, ..., a_n$ and $b_1, ..., b_n \in \mathbb{R}$,*

$$\inf_{\lambda} \left\{ (\lambda - 1)^2 + \sum_{i=1}^{n} (a_i - \lambda b_i)^2 \right\} \geq \frac{\sum_{i=1}^{n} (a_i - b_i)^2}{1 + \sum_{i=1}^{n} b_i^2}.$$

Proof.

$$(\lambda - 1)^2 + \sum_{i=1}^{n} (a_i - \lambda b_i)^2 = \lambda^2 \left(1 + \sum_{i=1}^{n} b_i^2 \right) - 2\lambda \left(1 + \sum_{i=1}^{n} a_i b_i \right) + \left(1 + \sum_{i=1}^{n} a_i^2 \right).$$

The minimum in x of $ax^2 - 2bx + c$, is $-\frac{b^2}{a} + c$, hence

$$\begin{aligned}
(\lambda - 1)^2 + \sum_{i=1}^{n} (a_i - \lambda b_i)^2 &\geq \left(1 + \sum_{i=1}^{n} a_i^2 \right) - \frac{\left(1 + \sum_{i=1}^{n} a_i b_i\right)^2}{(1 + \sum_{i=1}^{n} b_i^2)} \\
&= \frac{\sum_{i=1}^{n} (a_i - b_i)^2 - \left(\sum_{i=1}^{n} a_i b_i\right)^2 + \left(\sum_{i=1}^{n} a_i^2\right) \left(\sum_{i=1}^{n} b_i^2\right)}{1 + \sum_{i=1}^{n} b_i^2} \\
&\geq \frac{\sum_{i=1}^{n} (a_i - b_i)^2}{1 + \sum_{i=1}^{n} b_i^2}, \quad \text{using Cauchy-Schwartz inequality.}
\end{aligned}$$

□

Using lemma Appendix A.1, together with (1) and lemma Appendix C.1 which ensures that $\sum_{j \neq i} (R_{\theta,i,j})^2 \leq c < +\infty$ uniformly in i, θ and x, we obtain

$$\begin{aligned}
D_{\theta,\theta_0} &\geq ABC \frac{1}{1+c} \frac{1}{n} \sum_{i=1}^{n} \sum_{j \neq i} (R_{\theta,i,j} - R_{\theta_0,i,j})^2 \\
&= ABC \frac{1}{1+c} |R_\theta - R_{\theta_0}|^2, \quad \text{because } R_{\theta,i,i} = 1 = R_{\theta_0,i,i},
\end{aligned}$$

which proves (A.4) and ends the proof.

□

Appendix A.5. Proof of proposition 3.5

Proof. It is shown in [3] that $\frac{\partial}{\partial \theta_i} CV_\theta = \frac{2}{n} y^t M_\theta^i y = \frac{1}{n} y^t \left\{ M_\theta^i + \left(M_\theta^i\right)^t \right\} y$. We then show that $cov\left(y^t A y, y^t B y | X\right) = 2\mathrm{Tr}\left(\mathrm{AR}_{\theta_0} \mathrm{BR}_{\theta_0}\right)$ for symmetric matrices A and B, which shows (4).

A straightforward but relatively long calculation then shows

$$\begin{aligned}
\frac{\partial^2}{\partial\theta_i\partial\theta_j}CV_\theta &= -4\frac{1}{n}y^tR_\theta^{-1}\frac{\partial R_\theta}{\partial\theta_j}R_\theta^{-1}diag\left(R_\theta^{-1}\right)^{-3}diag\left(R_\theta^{-1}\frac{\partial R_\theta}{\partial\theta_i}R_\theta^{-1}\right)R_\theta^{-1}y \\
&\quad -4\frac{1}{n}y^tR_\theta^{-1}\frac{\partial R_\theta}{\partial\theta_i}R_\theta^{-1}diag\left(R_\theta^{-1}\right)^{-3}diag\left(R_\theta^{-1}\frac{\partial R_\theta}{\partial\theta_j}R_\theta^{-1}\right)R_\theta^{-1}y \\
&\quad +2\frac{1}{n}y^tR_\theta^{-1}\frac{\partial R_\theta}{\partial\theta_j}R_\theta^{-1}diag\left(R_\theta^{-1}\right)^{-2}R_\theta^{-1}\frac{\partial R_\theta}{\partial\theta_i}R_\theta^{-1}y \\
&\quad +6\frac{1}{n}y^tR_\theta^{-1}diag\left(R_\theta^{-1}\right)^{-4}diag\left(R_\theta^{-1}\frac{\partial R_\theta}{\partial\theta_j}R_\theta^{-1}\right)diag\left(R_\theta^{-1}\frac{\partial R_\theta}{\partial\theta_i}R_\theta^{-1}\right)R_\theta^{-1}y \\
&\quad -4\frac{1}{n}y^tR_\theta^{-1}diag\left(R_\theta^{-1}\right)^{-3}diag\left(R_\theta^{-1}\frac{\partial R_\theta}{\partial\theta_i}R_\theta^{-1}\frac{\partial R_\theta}{\partial\theta_j}R_\theta^{-1}\right)R_\theta^{-1}y \\
&\quad +2\frac{1}{n}y^tR_\theta^{-1}diag\left(R_\theta^{-1}\right)^{-3}diag\left(R_\theta^{-1}\frac{\partial^2R_\theta}{\partial\theta_i\partial\theta_j}R_\theta^{-1}\right)R_\theta^{-1}y \\
&\quad +2\frac{1}{n}y^tR_\theta^{-1}diag\left(R_\theta^{-1}\right)^{-2}R_\theta^{-1}\frac{\partial R_\theta}{\partial\theta_j}R_\theta^{-1}\frac{\partial R_\theta}{\partial\theta_i}R_\theta^{-1}y \\
&\quad +2\frac{1}{n}y^tR_\theta^{-1}diag\left(R_\theta^{-1}\right)^{-2}R_\theta^{-1}\frac{\partial R_\theta}{\partial\theta_i}R_\theta^{-1}\frac{\partial R_\theta}{\partial\theta_j}R_\theta^{-1}y \\
&\quad -2\frac{1}{n}y^tR_\theta^{-1}diag\left(R_\theta^{-1}\right)^{-2}R_\theta^{-1}\frac{\partial^2R_\theta}{\partial\theta_i\partial\theta_j}R_\theta^{-1}y.
\end{aligned}$$

We then have, using $\mathbb{E}\left(y^tAy|X\right) = \mathrm{Tr}\left(\mathrm{AR}_{\theta_0}\right)$ and for matrices D, M_1 and M_2, with D diagonal, $\mathrm{Tr}\left\{\mathrm{M_1Ddiag\left(M_2\right)}\right\} = \mathrm{Tr}\left\{\mathrm{M_2Ddiag\left(M_1\right)}\right\}$ and $\mathrm{Tr\left(DM_1\right) = Tr\left(DM_1^t\right)}$,

$$\begin{aligned}
\mathbb{E}\left(\frac{\partial^2}{\partial\theta_i\partial\theta_j}CV_{\theta_0}|X\right) &= -8\frac{1}{n}\mathrm{Tr}\left\{\frac{\partial \mathrm{R}_{\theta_0}}{\partial\theta_\mathrm{j}}\mathrm{R}_{\theta_0}^{-1}\mathrm{diag}\left(\mathrm{R}_{\theta_0}^{-1}\right)^{-3}\mathrm{diag}\left(\mathrm{R}_{\theta_0}^{-1}\frac{\partial \mathrm{R}_{\theta_0}}{\partial\theta_\mathrm{i}}\mathrm{R}_{\theta_0}^{-1}\right)\mathrm{R}_{\theta_0}^{-1}\right\} \qquad \text{(A.5)} \\
&\quad +2\frac{1}{n}\mathrm{Tr}\left\{\frac{\partial \mathrm{R}_{\theta_0}}{\partial\theta_\mathrm{j}}\mathrm{R}_{\theta_0}^{-1}\mathrm{diag}\left(\mathrm{R}_{\theta_0}^{-1}\right)^{-2}\mathrm{R}_{\theta_0}^{-1}\frac{\partial \mathrm{R}_{\theta_0}}{\partial\theta_\mathrm{i}}\mathrm{R}_{\theta_0}^{-1}\right\} \\
&\quad +6\frac{1}{n}\mathrm{Tr}\left\{\mathrm{diag}\left(\mathrm{R}_{\theta_0}^{-1}\right)^{-4}\mathrm{diag}\left(\mathrm{R}_{\theta_0}^{-1}\frac{\partial \mathrm{R}_{\theta_0}}{\partial\theta_\mathrm{j}}\mathrm{R}_{\theta_0}^{-1}\right)\mathrm{diag}\left(\mathrm{R}_{\theta_0}^{-1}\frac{\partial \mathrm{R}_{\theta_0}}{\partial\theta_\mathrm{i}}\mathrm{R}_{\theta_0}^{-1}\right)\mathrm{R}_{\theta_0}^{-1}\right\} \\
&\quad -4\frac{1}{n}\mathrm{Tr}\left\{\mathrm{diag}\left(\mathrm{R}_{\theta_0}^{-1}\right)^{-3}\mathrm{diag}\left(\mathrm{R}_{\theta_0}^{-1}\frac{\partial \mathrm{R}_{\theta_0}}{\partial\theta_\mathrm{i}}\mathrm{R}_{\theta_0}^{-1}\frac{\partial \mathrm{R}_{\theta_0}}{\partial\theta_\mathrm{j}}\mathrm{R}_{\theta_0}^{-1}\right)\mathrm{R}_{\theta_0}^{-1}\right\} \\
&\quad +2\frac{1}{n}\mathrm{Tr}\left\{\mathrm{diag}\left(\mathrm{R}_{\theta_0}^{-1}\right)^{-3}\mathrm{diag}\left(\mathrm{R}_{\theta_0}^{-1}\frac{\partial^2 \mathrm{R}_{\theta_0}}{\partial\theta_\mathrm{i}\partial\theta_\mathrm{j}}\mathrm{R}_{\theta_0}^{-1}\right)\mathrm{R}_{\theta_0}^{-1}\right\} \\
&\quad +4\frac{1}{n}\mathrm{Tr}\left\{\mathrm{diag}\left(\mathrm{R}_{\theta_0}^{-1}\right)^{-2}\mathrm{R}_{\theta_0}^{-1}\frac{\partial \mathrm{R}_{\theta_0}}{\partial\theta_\mathrm{i}}\mathrm{R}_{\theta_0}^{-1}\frac{\partial \mathrm{R}_{\theta_0}}{\partial\theta_\mathrm{j}}\mathrm{R}_{\theta_0}^{-1}\right\} \\
&\quad -2\frac{1}{n}\mathrm{Tr}\left\{\mathrm{diag}\left(\mathrm{R}_{\theta_0}^{-1}\right)^{-2}\mathrm{R}_{\theta_0}^{-1}\frac{\partial^2 \mathrm{R}_{\theta_0}}{\partial\theta_\mathrm{i}\partial\theta_\mathrm{j}}\mathrm{R}_{\theta_0}^{-1}\right\}.
\end{aligned}$$

The fourth and sixth terms of (A.5) are opposite and hence cancel each other. Indeed,

$$
\begin{aligned}
& \mathrm{Tr}\left\{ \mathrm{diag}\left(\mathrm{R}_{\theta_0}^{-1}\right)^{-3} \mathrm{diag}\left(\mathrm{R}_{\theta_0}^{-1} \frac{\partial \mathrm{R}_{\theta_0}}{\partial \theta_{\mathrm{i}}} \mathrm{R}_{\theta_0}^{-1} \frac{\partial \mathrm{R}_{\theta_0}}{\partial \theta_{\mathrm{j}}} \mathrm{R}_{\theta_0}^{-1} \right) \mathrm{R}_{\theta_0}^{-1} \right\} \\
= & \sum_{i=1}^{n} \left(R_{\theta_0}^{-1}\right)_{i,i}^{-3} \left(R_{\theta_0}^{-1} \frac{\partial R_{\theta_0}}{\partial \theta_i} R_{\theta_0}^{-1} \frac{\partial R_{\theta_0}}{\partial \theta_j} R_{\theta_0}^{-1} \right)_{i,i} \left(R_{\theta_0}^{-1}\right)_{i,i} \\
= & \sum_{i=1}^{n} \left(R_{\theta_0}^{-1}\right)_{i,i}^{-2} \left(R_{\theta_0}^{-1} \frac{\partial R_{\theta_0}}{\partial \theta_i} R_{\theta_0}^{-1} \frac{\partial R_{\theta_0}}{\partial \theta_j} R_{\theta_0}^{-1} \right)_{i,i} \\
= & \mathrm{Tr}\left\{ \mathrm{diag}\left(\mathrm{R}_{\theta_0}^{-1}\right)^{-2} \mathrm{R}_{\theta_0}^{-1} \frac{\partial \mathrm{R}_{\theta_0}}{\partial \theta_{\mathrm{i}}} \mathrm{R}_{\theta_0}^{-1} \frac{\partial \mathrm{R}_{\theta_0}}{\partial \theta_{\mathrm{j}}} \mathrm{R}_{\theta_0}^{-1} \right\}.
\end{aligned}
$$

Similarly the fifth and seventh terms of (A.5) cancel each other.

Hence, we show the expression of $\mathbb{E}\left(\frac{\partial^2}{\partial\theta_i\partial\theta_j} CV_{\theta_0}|X\right)$ of the proposition.

We use proposition Appendix C.7 to show the existence of $\Sigma_{CV,1}$ and $\Sigma_{CV,2}$. □

Appendix A.6. Proof of proposition 3.6

Proof. We use proposition Appendix C.9, with M^i the notation of proposition 3.5 and

$$
N_i = -\left\{ M^i + \left(M^i\right)^t \right\},
$$

together with propositions Appendix C.7 and 3.5 to show that

$$
\sqrt{n} \frac{\partial}{\partial \theta} CV_{\theta_0} \to \mathcal{N}\left(0, \Sigma_{CV,1}\right).
$$

We have seen in the proof of proposition 3.5 that there exist matrices $P_{i,j}$ in $\mathcal{M}_{\theta_0}$ (proposition Appendix C.7), so that $\frac{\partial^2}{\partial\theta_i\partial\theta_j} CV_{\theta_0} = \frac{1}{n} y^t P_{i,j} y$, with $\frac{1}{n}\mathrm{Tr}\left(\mathrm{P_{i,j}R}\right) \to \left(\Sigma_{\mathrm{CV},2}\right)_{\mathrm{i,j}}$ P_X-almost surely. Hence, using proposition Appendix C.8, $\frac{\partial^2}{\partial\theta^2} L_{\theta_0}$ converges to $\Sigma_{CV,2}$ in the mean square sense (on the product space).

Finally, $\frac{\partial^3}{\partial\theta_i\partial\theta_j\partial\theta_k} CV_{\tilde{\theta}}$ can be written as $\frac{1}{n}\left(z^t N_{\tilde{\theta}}^{i,j,k} z\right)$, where $N_{\tilde{\theta}}^{i,j,k}$ are sums of matrices of $\mathcal{M}_{\tilde{\theta}}$ (proposition Appendix C.7) and z depending on X and Y with $(z|X) = \mathcal{N}\left(0, I_n\right)$. The singular values of $N_{\tilde{\theta}}^{i,j,k}$ are bounded uniformly in $\tilde{\theta}$, n and x and so $\sup_{i,j,k,\tilde{\theta}} \left(\frac{\partial^3}{\partial\theta_i\partial\theta_j\partial\theta_k} CV_{\tilde{\theta}}\right)$ is bounded by $b\frac{1}{n} z^t z$, $b < +\infty$, and is hence bounded in probability. We apply proposition Appendix C.10 to conclude.

□

Appendix A.7. Proof of proposition 3.7

Proof. We show the proposition in the case $p = 1$, the generalization to the case $p > 1$ being the same as in proposition 3.3.

Similarly to the proof of proposition 3.1, we show that $\left| \frac{\partial^2}{\partial\theta^2} CV_{\theta_0} - \mathbb{E}\left(\frac{\partial^2}{\partial\theta^2} CV_{\theta_0}|X\right)\right| \to_p 0$. We will then show that there exists $C > 0$ so that P_X-a.s.,

$$
\mathbb{E}\left(\frac{\partial^2}{\partial\theta^2} CV_{\theta_0}|X\right) \geq C \left|\frac{\partial R_\theta}{\partial\theta}\right|^2. \tag{A.6}
$$

The proof of the proposition will hence be carried out similarly as in the proof of proposition 3.1.

$\frac{\partial^2}{\partial\theta^2} CV_{\theta_0}$ can be written as $z^t M z$ with z depending on X and Y and $(z|X) = \mathcal{N}\left(0, I_n\right)$, and M a sum of matrices of $\mathcal{M}_{\theta_0}$ (proposition Appendix C.7). Hence, using proposition Appendix

C.7, uniformly in n, $\sup_\theta \left|\frac{\partial^2}{\partial\theta^2} CV_\theta\right| \leq a\frac{1}{n}z^t z$ with $a < +\infty$. Hence, for fixed n, we can exchange derivatives and means conditionally to X and so

$$\mathbb{E}\left(\frac{\partial^2}{\partial\theta^2} CV_{\theta_0}|X\right) = \frac{\partial^2}{\partial\theta^2}\mathbb{E}\left(CV_{\theta_0}|X\right).$$

Then, with $r_{i,\theta}$, $R_{-i,\theta}$ and y_{-i} the notation of the proof of proposition 3.4,

$$\begin{aligned}\mathbb{E}\left(CV_\theta|X\right) &= \frac{1}{n}\sum_{i=1}^n\left[1 - r_{i,\theta_0}^t R_{-i,\theta_0}^{-1} r_{i,\theta_0} + \mathbb{E}\left\{\left(r_{i,\theta_0}^t R_{-i,\theta_0}^{-1} y_{-i} - r_{i,\theta}^t R_{-i,\theta}^{-1} y_{-i}\right)^2 |X\right\}\right] \\ &= \frac{1}{n}\sum_{i=1}^n\left(1 - r_{i,\theta_0}^t R_{-i,\theta_0}^{-1} r_{i,\theta_0}\right) + \frac{1}{n}\sum_{i=1}^n\left(r_{i,\theta}^t R_{-i,\theta}^{-1} - r_{i,\theta_0}^t R_{-i,\theta_0}^{-1}\right) R_{-i,\theta_0}\left(R_{-i,\theta}^{-1} r_{i,\theta} - R_{-i,\theta_0}^{-1} r_{i,\theta_0}\right).\end{aligned}$$

By differentiating twice with respect to θ and taking the value at θ_0 we obtain

$$\begin{aligned}\mathbb{E}\left(\frac{\partial^2}{\partial\theta^2} CV_{\theta_0}|X\right) &= \frac{1}{n}\sum_{i=1}^n\left\{\frac{\partial}{\partial\theta}\left(R_{-i,\theta_0}^{-1} r_{i,\theta_0}^t\right)\right\}^t R_{-i,\theta_0}\left\{\frac{\partial}{\partial\theta}\left(R_{-i,\theta_0}^{-1} r_{i,\theta_0}^t\right)\right\} \\ &\geq A\frac{1}{n}\sum_{i=1}^n\left|\left|\left\{\frac{\partial}{\partial\theta}\left(R_{-i,\theta_0}^{-1} r_{i,\theta_0}^t\right)\right\}\right|\right|^2 \quad \text{with } A = \inf_{n,i,x}\phi_i^2\left(R_{-i,\theta_0}\right),\ A > 0,\end{aligned}$$

then, using the virtual CV formulas [18, 7],

$$\begin{aligned}\mathbb{E}\left(\frac{\partial^2}{\partial\theta^2} CV_{\theta_0}|X\right) &\geq A\frac{1}{n}\sum_{i=1}^n\sum_{j\neq i}\left[\frac{\partial}{\partial\theta}\left\{\frac{\left(R_{\theta_0}^{-1}\right)_{i,j}}{\left(R_{\theta_0}^{-1}\right)_{i,i}}\right\}\right]^2 \\ &= A\left|\frac{\partial}{\partial\theta}\left\{diag\left(R_{\theta_0}^{-1}\right)^{-1} R_{\theta_0}^{-1}\right\}\right|^2 \\ &= A\left|diag\left(R_{\theta_0}^{-1}\right)^{-1} diag\left(R_{\theta_0}^{-1}\frac{\partial R_{\theta_0}}{\partial\theta}R_{\theta_0}^{-1}\right) diag\left(R_{\theta_0}^{-1}\right)^{-1} R_{\theta_0}^{-1} - diag\left(R_{\theta_0}^{-1}\right)^{-1}\left(R_{\theta_0}^{-1}\frac{\partial R_{\theta_0}}{\partial\theta}R_{\theta_0}^{-1}\right)\right|^2 \\ &\geq A^2 B\left|diag\left(R_{\theta_0}^{-1}\frac{\partial R_{\theta_0}}{\partial\theta}R_{\theta_0}^{-1}\right) diag\left(R_{\theta_0}^{-1}\right)^{-1} - R_{\theta_0}^{-1}\frac{\partial R_{\theta_0}}{\partial\theta}\right|^2 \quad \text{with } B = \inf_{i,n,x}\phi_i\left(R_{\theta_0}^{-1}\right),\ B > 0 \\ &\geq A^2 B\inf_{\lambda_1,\ldots,\lambda_n}\left|D_\lambda - R_{\theta_0}^{-1}\frac{\partial R_{\theta_0}}{\partial\theta}\right|^2 \\ &\geq A^2 B^2\inf_{\lambda_1,\ldots,\lambda_n}\left|R_{\theta_0}D_\lambda - \frac{\partial R_{\theta_0}}{\partial\theta}\right|^2.\end{aligned}$$

Then, as $K_\theta(0) = 1$ for all θ, and hence $\frac{\partial}{\partial\theta}K_{\theta_0}(0) = 0$,

$$\begin{aligned}\mathbb{E}\left(\frac{\partial^2}{\partial\theta^2} CV_{\theta_0}|X\right) &\geq A^2 B^2\inf_{\lambda_1,\ldots,\lambda_n}\frac{1}{n}\sum_{i=1}^n\left[\lambda_i^2 + \sum_{j\neq i}\left\{\lambda_i\left(R_{\theta_0}\right)_{i,j} - \left(\frac{\partial R_{\theta_0}}{\partial\theta}\right)_{i,j}\right\}^2\right] \\ &= A^2 B^2\frac{1}{n}\sum_{i=1}^n\inf_\lambda\left[\lambda^2 + \sum_{j\neq i}\left\{\lambda\left(R_{\theta_0}\right)_{i,j} - \left(\frac{\partial R_{\theta_0}}{\partial\theta}\right)_{i,j}\right\}^2\right].\end{aligned}$$

We then show, similarly to lemma Appendix A.1, that

$$\lambda^2 + \sum_{i=1}^n\left(a_i - \lambda b_i\right)^2 \geq \frac{\sum_{i=1}^n a_i^2}{1 + \sum_{i=1}^n b_i^2}. \tag{A.7}$$

Hence, with $C \in [1, +\infty)$, by using (1) and lemma Appendix C.1,

$$\begin{aligned}\mathbb{E}\left(\frac{\partial^2}{\partial\theta^2} C V_{\theta_0} | X\right) &\geq \frac{A^2B^2}{C}\frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}\left\{\left(\frac{\partial R_{\theta_0}}{\partial\theta}\right)_{i,j}\right\}^2 \\ &= \frac{A^2B^2}{C}\left|\frac{\partial R_{\theta_0}}{\partial\theta}\right|^2 \quad \text{because } \tfrac{\partial}{\partial\theta}K_{\theta_0(0)} = 0.\end{aligned}$$

We then showed (A.6), which concludes the proof in the case $p = 1$.

□

Appendix B. Proofs for section 4

Appendix B.1. Proof of proposition 4.2

Proof. It is enough to show the proposition for $\epsilon \in [0, \alpha]$ for all $\alpha < \frac{1}{2}$. We use the following lemma.

Lemma Appendix B.1. *Let f_n be a sequence of C^2 functions on a segment of $\mathbb{R}$. We assume $f_n \to_{unif} f$, $f_n' \to_{unif} g$, $f_n'' \to_{unif} h$. Then, f is C^2, $f' = g$, and $f'' = h$.*

We denote $f_n(\epsilon) = \frac{1}{n}\mathbb{E}\{\mathrm{Tr}(\mathrm{M_n})\}$ where $(M_n)_{n\in\mathbb{N}^*}$ is a random matrix sequence defined on $(\Omega_X, \mathcal{F}_X, P_X)$ which belongs to $\mathcal{M}_\theta$ (proposition Appendix C.7). We showed that f_n converges simply to Σ on $[0, \alpha]$. We firstly use the dominated convergence theorem to show that f_n is C^2 and that f_n' and f_n'' are of the form

$$\mathbb{E}\left\{\frac{1}{n}\mathrm{Tr}(\mathrm{N})\right\}, \tag{B.1}$$

with N a sum of random matrix sequences of $\tilde{\mathcal{M}}_{\theta_0}$. $\tilde{\mathcal{M}}_{\theta_0}$ is similar to $\mathcal{M}_{\theta_0}$ (proposition Appendix C.7), with the addition of the derivative matrices with respect to ϵ. We can then, using (6), adapt proposition Appendix C.7 to show that f_n' and f_n'' converge simply to some functions g and h on $[0, \alpha]$.

Finally, still adapting proposition Appendix C.7, the singular values of N are bounded uniformly in x and n. Hence, using $\mathrm{Tr}(\mathrm{M}) \leq \mathrm{n}||\mathrm{M}||$, the derivatives of f_n, f_n' and f_n'' are bounded uniformly in n, so that the simple convergence implies the uniform convergence. The conditions of lemma Appendix B.1 are hence fulfilled.

□

Appendix C. Technical results

In subsection Appendix C.1 we state several technical results that are used in the proofs of the results of sections 3 and 4. Proofs are given in subsection Appendix C.2.

Appendix C.1. Statement of the technical results

Lemma Appendix C.1. *Let $f : \mathbb{R}^d \to \mathbb{R}^+$, so that $f(t) \leq \frac{1}{1+|t|_\infty^{d+1}}$. Then, for all $i \in \mathbb{N}^*$, $\epsilon \in (-\frac{1}{2}, \frac{1}{2})$ and $(x_i)_{i\in\mathbb{N}^*} \in S_X^{\mathbb{N}^*}$,*

$$\sum_{j\in\mathbb{N}^*, j\neq i} f\{v_i - v_j + \epsilon(x_i - x_j)\} \leq 2^d d \sum_{j\in\mathbb{N}} \frac{\left(j + \frac{3}{2}\right)^{d-1}}{1 + j^{d+1}}.$$

Lemma Appendix C.2. *Let $f : \mathbb{R}^d \to \mathbb{R}^+$, so that $f(t) \leq \frac{1}{1+|t|_\infty^{d+1}}$. We consider $\delta < \frac{1}{2}$. Then, for all $i \in \mathbb{N}^*$, $a > 0$, $\epsilon \in [-\delta, \delta]$ and $(x_i)_{i\in\mathbb{N}^*} \in S_X^{\mathbb{N}^*}$,*

$$\sum_{j\in\mathbb{N}^*, j\neq i} f[a\{v_i - v_j + \epsilon(x_i - x_j)\}] \leq 2^d d \sum_{j\in\mathbb{N}} \frac{\left(j + \frac{3}{2}\right)^{d-1}}{1 + a^{d+1}(j + 1 - 2\delta)^{d+1}}.$$

Proof. Similar to the proof of lemma Appendix C.1. □

Lemma Appendix C.3. *Let $f : \mathbb{R}^d \to \mathbb{R}^+$, so that $f(t) \leq \frac{1}{1+|t|_\infty^{d+1}}$. Then, for all $i \in \mathbb{N}^*$, $N \in \mathbb{N}^*$ and $(x_i)_{i\in\mathbb{N}^*} \in S_X^{\mathbb{N}^*}$,*

$$\sum_{j\in\mathbb{N}^*, |v_i - v_j|_\infty \geq N} f\{v_i - v_j + \epsilon(x_i - x_j)\} \leq 2^d d \sum_{j\in\mathbb{N}, j\geq N-1} \frac{\left(j+\frac{3}{2}\right)^{d-1}}{1+j^{d+1}}.$$

Proof. Similar to the proof of lemma Appendix C.1. □

Proposition Appendix C.4. *Assume that condition 2.1 is satisfied.*

For all $0 \leq \delta < \frac{1}{2}$, there exists $C_\delta > 0$ so that for all $|\epsilon| \leq \delta$, for all $\theta \in \Theta$, for all $n \in \mathbb{N}^*$ and for all $x \in (S_X)^n$, the eigenvalues of R_θ are larger than C_δ.

Lemma Appendix C.5. *Assume that condition 2.1 is satisfied.*

For all $|\epsilon| < \frac{1}{2}$ and for all $K \in \mathbb{N}$, there exists $C_{\epsilon,K}$ so that the eigenvalues of R_θ^{-1} and of $\frac{\partial^q R_\theta}{\partial\theta_{i_1},...,\partial\theta_{i_q}}$, $0 \leq q \leq K$, $1 \leq i_1, ..., i_q \leq p$, are bounded by $C_{\epsilon,K}$, uniformly in $n \in \mathbb{N}$, $x \in (S_X)^n$ and $\theta \in \Theta$.

Proof. Using, proposition Appendix C.4, we control the eigenvalues of R_θ^{-1} uniformly in x and θ.

With (1) and lemma Appendix C.1, and using Gershgorin circle theorem, we control the eigenvalues of $\frac{\partial^q R_\theta}{\partial\theta_{i_1},...,\partial\theta_{i_q}}$.

□

Lemma Appendix C.6. *For M symmetric real non-negative matrix, $\inf_i \phi_i(diag(M)) \geq \inf_i \phi_i(M)$ and $\sup_i \phi_i(diag(M)) \leq \sup_i \phi_i(M)$. Furthermore, if for two sequences of symmetric matrices M_n and N_n, $M_n \sim N_n$, then $diag(M_n) \sim diag(N_n)$.*

Proof. We use $M_{i,i} = e_i^t M e_i$, where $(e_i)_{i=1...n}$ is the standard basis of $\mathbb{R}^n$. Hence $\inf_i \phi_i(M) \leq M_{i,i} \leq \sup_i \phi_i(M)$ for a symmetric real non-negative matrix M. We also use $|diag(M)| \leq |M|$. □

The next proposition gives a law of large numbers for the matrices that can be written using only matrix multiplications, the matrix R_θ^{-1}, the matrices $\frac{\partial^k}{\partial\theta_1,...,\partial\theta_k} R_\theta$, the *diag* operator applied to the symmetric products of matrices R_θ, R_θ^{-1} and $\frac{\partial^k}{\partial\theta_1,...,\partial\theta_k} R_\theta$, and the matrix $diag\left(R_\theta^{-1}\right)^{-1}$. Examples of sums of these matrices are the matrices Σ_{ML}, $\Sigma_{CV,1}$ and $\Sigma_{CV,2}$ of propositions 3.2 and 3.5.

Proposition Appendix C.7. *Assume that condition 2.1 is satisfied.*

Let $\theta \in \Theta$. We denote the set of multi-indexes $S_p := \cup_{k\in\{0,1,2,3\}} \{1, ..., p\}^k$. For $I = (i_1, ..., i_k) \in S_p$, we denote $n(I) = k$. Then, we denote for $I \in S_p \cup \{-1\}$,

$$R_\theta^I := \begin{cases} \frac{\partial^{n(I)}}{\partial\theta_{I_1},...,\partial\theta_{I_{n(I)}}} R_\theta & \text{if } I \in S_p \\ R_\theta^{-1} & \text{if } I = -1 \end{cases}.$$

We then denote

- $M_{nd}^I = R_\theta^I \qquad$ *for* $I \in S_{nd} := (S_p \cup \{-1\})$
- $M_{sd}^1 = diag\left(R_\theta^{-1}\right)^{-1}$
- $M_{bd}^I = diag\left(R_\theta^{I_1} ... R_\theta^{I_{n(I)}}\right) \qquad$ *for* $I \in S_{bd} := \cup_{k\in\mathbb{N}^*} S_{nd}^k$

We then define $\mathcal{M}_\theta$ as the set of sequences of random matrices (defined on $(\Omega_X, \mathcal{F}_X, P_X)$), indexed by $n \in \mathbb{N}^*$, dependent on X, which can be written $M_{d_1}^{I_1} ... M_{d_K}^{I_K}$ with $\{d_1, I_1\}, ..., \{d_K, I_K\} \in (\{nd\} \times S_{nd}) \cup (\{sd\} \times \{1\}) \cup (\{bd\} \times S_{bd})$, and so that, for the matrices $M_{d_j}^{I_j}$, so that $d_j = bd$, the matrix $R_\theta^{(I_j)_1} ... R_\theta^{(I_j)_{n(I_j)}}$ be symmetric.

Then, for every matrix $M_{d_1}^{I_1} ... M_{d_K}^{I_K}$ of $\mathcal{M}_\theta$, the singular values of $M_{d_1}^{I_1} ... M_{d_K}^{I_K}$ are bounded uniformly in θ, n and $x \in (S_X)^n$. Then, denoting $S_n := \frac{1}{n}\mathrm{Tr}\left(\mathrm{M_{d_1}^{I_1} ... M_{d_K}^{I_K}}\right)$, there exists a deterministic limit S, which only depends on ϵ, θ and $(d_1, I_1), ..., (d_K, I_K)$, so that $S_n \to S$ P_X-almost surely. Hence $S_n \to S$ in quadratic mean and $var(S_n) \to 0$ as $n \to +\infty$.

Proposition Appendix C.8. *Assume that condition 2.1 is satisfied.*

Let $M \in \mathcal{M}_\theta$ (proposition Appendix C.7). Then, $\frac{1}{n} y^t M y$ converges to $\Sigma := \lim_{n \to +\infty} \frac{1}{n}\mathrm{Tr}\left(\mathrm{MR}_{\theta_0}\right)$, in the mean square sense (on the product space).

Proposition Appendix C.9. *Assume that condition 2.1 is satisfied.*

We recall $X \sim \mathcal{L}_X^{\otimes n}$ and $y_i = Y(i + \epsilon X_i)$, $1 \leq i \leq n$. We consider symmetric matrix sequences $M_1, ..., M_p$ and $N_1, ..., N_p$ (defined on $(\Omega_X, \mathcal{F}_X, P_X)$), functions of X, so that the eigenvalues of $N_1, ..., N_p$ are bounded uniformly in n and $x \in (S_X)^n$, $\mathrm{Tr}\left(\mathrm{M_i + N_i R}\right) = 0$ for $1 \leq i \leq p$ and there exists a $p \times p$ matrix Σ so that $\frac{1}{n}\mathrm{Tr}\left(\mathrm{N_i R N_j R}\right) \to (\Sigma)_{\mathrm{i,j}}$ P_X-almost surely. Then the sequence of p-dimensional random vectors (defined on the product space) $\left(\frac{1}{\sqrt{n}}\left\{\mathrm{Tr}\left(\mathrm{M_i}\right) + \mathrm{y^t N_i y}\right\}\right)_{i=1...p}$ converges in law to a Gaussian random vector with mean zero and covariance matrix 2Σ.

Proposition Appendix C.10. *We recall $X \sim \mathcal{L}_X^{\otimes n}$ and $y_i = Y(i + \epsilon X_i)$, $1 \leq i \leq n$. We consider a consistent estimator $\hat{\theta} \in \mathbb{R}^p$ so that $\mathbb{P}\left(c\left(\hat{\theta}\right) = 0\right) \to 1$, for a function $c : \Theta \to \mathbb{R}^p$, dependent on X and Y, and twice differentiable in θ. We assume that $\sqrt{n} c(\theta_0) \to \mathcal{N}(0, \Sigma_1)$, for a $p \times p$ matrix Σ_1 ant that the matrix $\frac{\partial c(\theta_0)}{\partial \theta}$ converges in probability to a $p \times p$ positive matrix Σ_2 (convergences are defined on the product space). Finally we assume that $\sup_{\tilde{\theta}, i, j, k} \left|\frac{\partial^2}{\partial \theta_i \partial \theta_j} c_k\left(\tilde{\theta}\right)\right|$ is bounded in probability.*

Then

$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) \to \mathcal{N}\left(0, \Sigma_2^{-1}\Sigma_1\Sigma_2^{-1}\right).$$

Proposition Appendix C.10 can be proved using standard M-estimator techniques. In subsection Appendix C.2 we give a short proof for consistency.

Appendix C.2. Proof of the technical results

Proof of lemma Appendix C.1.

Proof.

$$\begin{aligned}\sum_{j \in \mathbb{N}, j \neq i} f\left\{v_i - v_j + \epsilon\left(x_i - x_j\right)\right\} &\leq \sum_{v \in \mathbb{Z}^d, v \neq 0} \sup_{\delta_v \in [-1,1]^d} f\left(v + \delta_v\right) \\ &= \sum_{j \in \mathbb{N}} \sum_{v \in \{-j-1, ..., j+1\}^d \backslash \{-j, j\}^d} \sup_{\delta_v \in [-1,1]^d} f\left(v + \delta_v\right).\end{aligned}$$

For $v \in \{-j-1, ..., j+1\}^d \backslash \{-j, j\}^d$, $|v + \delta_v|_\infty \geq j$. The cardinality of the set $\{-j-1, ..., j+1\}^d \backslash \{-j, j\}^d$ is

$$(2j+3)^d - (2j+1)^d = \int_{2j+1}^{2j+3} d.t^{d-1} dt \leq 2d(2j+3)^{d-1} = 2^d d\left(j + \frac{3}{2}\right)^{d-1}.$$

Hence

$$\sum_{j \in \mathbb{N}} f\left\{v_i - v_j + \epsilon\left(x_i - x_j\right)\right\} \leq \sum_{j \in \mathbb{Z}} 2^d d\left(j + \frac{3}{2}\right)^{d-1} \frac{1}{1 + j^{d+1}}.$$

□

Proof of proposition Appendix C.4.

Proof. Let $h : \mathbb{R}^d \to \mathbb{R}$ so that $\hat{h}(f) = \mathbf{1}_{|f|^2 \in [-1,1]} \exp\left(-\frac{1}{|f|^2-1|}\right)$. Then, $\hat{h}$ is C^∞ and with compact support, so there exists $C > 0$ so that $|h(t)| \leq \frac{C}{1+|t|_\infty^{d+1}}$.

Hence, from lemma Appendix C.2, there exists $0 < a < \infty$ so that for all $i \in \mathbb{N}$,

$$\sum_{j \in \mathbb{N}, j \neq i} h\left[a\left\{v_i - v_j + \epsilon\left(x_i - x_j\right)\right\}\right] \leq C 2^d d \sum_{j \in \mathbb{Z}} \frac{(j+1)^{d-1}}{1 + a^{d+1}\left(j + 1 - 2\delta\right)^{d+1}} \leq \frac{1}{2} h(0).$$

Hence, using Gershgorin circle theorem, for all $t_1, ..., t_n \in \mathbb{R}$, $x_1, ..., x_n \in S_X$,

$$\begin{aligned}
\frac{1}{2} h(0) \sum_{i=1}^n t_i^2 &\leq \sum_{i,j=1}^n t_i t_j h\left[a\left\{v_i - v_j + \epsilon\left(x_i - x_j\right)\right\}\right] \\
&= \sum_{i,j=1}^n t_i t_j \frac{1}{a} \int_{\mathbb{R}^d} \hat{h}\left(\frac{f}{a}\right) e^{\mathrm{i} f \cdot \left\{v_i - v_j + \epsilon\left(x_i - x_j\right)\right\}} df \\
&= \frac{1}{a} \int_{\mathbb{R}^d} \hat{h}\left(\frac{f}{a}\right) \left| \sum_{i=1}^n t_i e^{\mathrm{i} f \cdot \left(v_i + \epsilon x_i\right)} \right|^2 df.
\end{aligned}$$

Hence, as $(\theta, f) \to \hat{K}_\theta(f)$ is continuous and positive, using a compacity argument, there exists $C_2 > 0$ so that for all $\theta \in \Theta$, $f \in [-a, a]^d$, $\hat{K}_\theta(f) \geq C_2 \hat{h}\left(\frac{f}{a}\right)$. Hence,

$$\begin{aligned}
\frac{1}{2} h(0) \sum_{i=1}^n Y_i^2 &\leq \frac{1}{a C_2} \int_{\mathbb{R}^d} \hat{K}_\theta(f) \left| \sum_{i=1}^n t_i e^{\mathrm{i} f \cdot \left(v_i + \epsilon x_i\right)} \right|^2 df, \\
&= \frac{1}{a C_2} \sum_{i,j=1}^n t_i t_j K_\theta\left\{v_i - v_j + \epsilon\left(x_i - x_j\right)\right\}.
\end{aligned}$$

□

Proof of proposition Appendix C.7.

Proof. Let $M_{d_1}^{I_1} ... M_{d_K}^{I_K} \in \mathcal{M}_\theta$ be fixed in the proof.

The eigenvalues of R_θ^I, $I \in S_{nd}$, are bounded uniformly with respect to n, θ and x (lemma Appendix C.5). Then, using lemma Appendix C.6, we show that the eigenvalues of $diag\left(R_\theta^{-1}\right)^{-1}$ are bounded uniformly in x, n and θ. Then, for $M_{bd}^I = diag\left(R_\theta^{I_1} ... R_\theta^{I_{n(I)}}\right)$, the eigenvalues of $R_\theta^{I_1} ... R_\theta^{I_{n(I)}}$ are bounded by the product of the eigenvalues of $R_\theta^{I_1}, ..., R_\theta^{I_{n(I)}}$. Hence we use lemma Appendix C.6 to show that the eigenvalues of M_{bd}^I are bounded uniformly in n, θ and x. Finally we use $||A_1 ... A_K|| \leq ||A_1|| ... ||A_K||$ to show that $||M_{d_1}^{I_1} ... M_{d_K}^{I_K}||$ is bounded uniformly in n, θ and x.

We decompose n into $n = N_1^d n_2 + r$ with $N_1, n_2, r \in \mathbb{N}$ and $r < N_1^d$. We define $C(v_i)$ as the unique $v \in \mathbb{N}^d$ so that $v_i \in \prod_{k=1}^d \{N_1 v_k + 1, ..., N_1(v_k + 1)\}$.

We then define the sequence of matrices $\tilde{R}_\theta$ by $\left(\tilde{R}_\theta\right)_{i,j} = (R_\theta)_{i,j} \mathbf{1}_{C(v_i) = C(v_j)}$. We denote $\tilde{M}_{d_1}^{I_1} ... \tilde{M}_{d_K}^{I_K}$ the matrix built by replacing R_θ by $\tilde{R}_\theta$ in the expression of $M_{d_1}^{I_1} ... M_{d_K}^{I_K}$ (we also make the substitution for the inverse and the partial derivatives).

Lemma Appendix C.11. $\left|\tilde{M}_{d_1}^{I_1} ... \tilde{M}_{d_K}^{I_K} - M_{d_1}^{I_1} ... M_{d_K}^{I_K}\right|^2 \to 0$, *uniformly in* $x \in (S_X)^n$, *when* $N_1, n_2 \to \infty$.

Proof. Let $\delta > 0$ and N so that $T_N := C_0^2 2^{2d} d^2 \sum_{j\in\mathbb{N}, j\geq N-1} \frac{\left(j+\frac{3}{2}\right)^{2(d-1)}}{(1+j^{d+1})^2} \leq \delta$. Then:

$$\begin{aligned}
\left|\tilde{R}_\theta - R_\theta\right|^2 &= \frac{1}{n}\sum_{i,j=1}^{n}\left\{(R_\theta)_{i,j} - \left(\tilde{R}_\theta\right)_{i,j}\right\}^2, \\
&= \frac{1}{n}\sum_{1\leq i,j\leq n, C(v_i)\neq C(v_j)} K_\theta^2\left\{v_i - v_j + \epsilon\left(x_i - x_j\right)\right\}, \\
&\leq \frac{1}{n}\sum_{i=1}^{n}\sum_{j\in\mathbb{N}^*, C(v_i)\neq C(v_j)} K_\theta^2\left\{v_i - v_j + \epsilon\left(x_i - x_j\right)\right\}.
\end{aligned}$$

There exists a unique a so that $(aN_1)^d \leq n < \{(a+1)N_1\}^d$. Among the n observation points, $(aN_1)^d$ are in the $C(v)$, for $v \in \{1,...,a\}^d$. The number of remaining points is less than $dN_1\{(a+1)N_1\}^{d-1}$. Therefore, using (1),

$$\begin{aligned}
\left|\tilde{R}_\theta - R_\theta\right|^2 &\leq \frac{1}{n}\sum_{v\in\{1,...,a\}^d}\sum_{1\leq i\leq n, v_i\in C(v)}\sum_{j\in\mathbb{N}^*, C(v_j)\neq C(v_i)} K_\theta^2\left\{v_i - v_j + \epsilon\left(x_i - x_j\right)\right\} + \frac{1}{n}dN_1\left\{(a+1)N_1\right\}^{d-1}T_0, \\
&= \frac{1}{n}\sum_{v\in\{1,...,a\}^d}\sum_{1\leq i\leq n, v_i\in C(v)}\sum_{j\in\mathbb{N}^*, C(v_j)\neq C(v_i)} K_\theta^2\left\{v_i - v_j + \epsilon\left(x_i - x_j\right)\right\} + o(1).
\end{aligned}$$

Then, for fixed v, the cardinality of the set of the integers $i \in \{1,...,n\}$, so that $v_i \in C(v)$ and there exists $j \in \mathbb{N}^*$ so that $|v_i - v_j|_\infty \leq N$ is $N_1^d - (N_1 - 2N)^d$ and is less than $2NdN_1^{d-1}$. Hence, using (1), lemmas Appendix C.1 and Appendix C.3,

$$\begin{aligned}
\left|\tilde{R}_\theta - R_\theta\right|^2 &\leq \frac{1}{n}\sum_{v\in\{1,...,a\}^d}\left(2NdN_1^{d-1}T_0 + N_1^d T_N\right) + o(1) \\
&\leq \frac{1}{a^d N_1^d}a^d\left\{\left(2NdN_1^{d-1}T_0 + N_1^d T_N\right)\right\} + o(1).
\end{aligned}$$

This last term is smaller than 2δ for N_1 and n_2 large enough. Hence we showed $\left|\tilde{R}_\theta - R_\theta\right| \to 0$ uniformly in x, when $N_1, n_2 \to \infty$. We can show the same result for $\frac{\partial^k R_\theta}{\partial\theta_1...\partial\theta_k}$ and $\frac{\partial^k \tilde{R}_\theta}{\partial\theta_1...\partial\theta_k}$. Finally we use [8] theorem 2.1 to show that $\left|\tilde{R}_\theta^{-1} - R_\theta^{-1}\right| \to 0$ uniformly in x, when $N_1, n_2 \to \infty$. Hence, using [8], theorem 2.1 and lemma Appendix C.6, $|\tilde{M}_d^I - M_d^I|$ converges to 0 uniformly in x when $N_1, n_2 \to \infty$, for $d \in \{nd, sd, bd\}$ and $I \in S_{nd} \cup \{1\} \cup S_{bd}$. We conclude using [8], theorem 2.1.

□

We denote, for every N_1, n_2 and r, with $0 \leq r < N_1^d$, $n = N_1^d n_2 + r$ and $S_{N_1,n_2} := \frac{1}{n}\mathrm{Tr}\left(\tilde{\mathrm{M}}_{\mathrm{d}_1}^{\mathrm{I}_1}...\tilde{\mathrm{M}}_{\mathrm{d}_\mathrm{K}}^{\mathrm{I}_\mathrm{K}}\right)$, which is a sequence of real random variables defined on $(\Omega_X, \mathcal{F}_X, P_X)$ and indexed by N_1, n_2 and r. Using [8], corollary 2.1 and lemma Appendix C.11, $|S_n - S_{N_1,n_2}| \to 0$ uniformly in x when $N_1, n_2 \to \infty$ (uniformly in r). As the matrices in the expression of S_{N_1,n_2} are block diagonal, we can write $S_{N_1,n_2} = \frac{1}{n_2}\sum_{l=1}^{n_2} S_{N_1^d}^l + o\left(\frac{1}{n_2}\right)$, where the $S_{N_1^d}^l$ are iid random variables defined on $(\Omega_X, \mathcal{F}_X, P_X)$ with the distribution of $S_{N_1^d}$. We denote $\bar{S}_{N_1^d} := E_X\left(S_{N_1^d}\right)$. Then, using the strong law of large numbers, for fixed N_1, $S_{N_1,n_2} \to \bar{S}_{N_1^d}$ P_X-almost surely when $n_2 \to \infty$ (uniformly in r).

For every $N_1, p_{N_1}, n_2 \in \mathbb{N}^*$, there exist a unique $n_2', r \in \mathbb{N}^*$ so that $(N_1 + p_{N_1})^d n_2 = N_1^d n_2' + r$.

Then we have

$$\begin{aligned}|\bar{S}_{(N_1)^d} - \bar{S}_{\left(N_1+p_{N_1}\right)^d}| &\leq |\bar{S}_{(N_1)^d} - S_{N_1,n_2'}| + |S_{N_1,n_2'} - S_{N_1^d n_2'+r}| \\ &\quad + |S_{N_1^d n_2'+r} - S_{\left(N_1+p_{N_1}\right)^d n_2}| + |S_{\left(N_1+p_{N_1}\right)^d n_2} - S_{N_1+p_{N_1},n_2}| + |S_{N_1+p_{N_1},n_2} - \bar{S}_{\left(N_1+p_{N_1}\right)^d}| \\ &= A + B + C + D + E. \end{aligned} \tag{C.1}$$

Because n_2' and r depend on N_1, p_{N_1} and n_2, A, B, C, D and E are sequences of random variables defined on $(\Omega_X, \mathcal{F}_X, P_X)$ and indexed by N_1, p_{N_1} and n_2. We have seen that there exists $\tilde{\Omega}_X \subset \Omega_X$, with $P_X(\tilde{\Omega}_X) = 1$ so that for $\omega_X \in \tilde{\Omega}_X$, when $N_1, n_2 \to +\infty$, we also have $N_1 + p_{N_1}, n_2' \to +\infty$, and so B and D converge to zero.

Now, for every $N_1 \in \mathbb{N}^*$, let Ω_{X,N_1} be so that $P_X(\Omega_{X,N_1}) = 1$ and for all $\omega_X \in \Omega_{X,N_1}$, $S_{N_1,n_2} \to_{n_2 \to +\infty} \bar{S}_{N_1^d}$. Let $\tilde{\tilde{\Omega}}_X = \cap_{N_1 \in \mathbb{N}^*} \Omega_{X,N_1}$. Then $P_X(\tilde{\tilde{\Omega}}_X) = 1$ and for all $\omega_X \in \tilde{\tilde{\Omega}}_X$, for all $N_1 \in \mathbb{N}^*$, $S_{N_1,n_2} \to_{n_2 \to +\infty} \bar{S}_{N_1^d}$.

We will now show that $N_1 \to \bar{S}_{N_1^d}$ is a Cauchy sequence. Let $\delta > 0$. $P_X(\tilde{\tilde{\Omega}} \cap \tilde{\Omega}) = 1$ so this set in non-empty. Let us fix $\omega_X \in \tilde{\tilde{\Omega}} \cap \tilde{\Omega}$. In (C.1), C is null. There exist $\bar{N}_1$ and $\bar{N}_2$ so that for every $N_1 \geq \bar{N}_1$, $n_2 \geq \bar{n}_2$, $p_{N_1} > 0$, B and D are smaller than δ. Let us now fix any $N_1 \geq \bar{N}_1$. Then, for every $p_{N_1} > 0$, with $n_2 \geq \bar{n}_2$ large enough, A and E are smaller than δ.

Hence, we showed that $N_1 \to \bar{S}_{(N_1)^d}$ is a Cauchy sequence and we denote its limit by S. Since $N_1 \to \bar{S}_{(N_1)^d}$ is deterministic, S is deterministic and $\bar{S}_{(N_1)^d} \to_{N_1 \to +\infty} S$.

Finally, let $n = N_1^d n_2 + r$ with $N_1, n_2 \to \infty$. Then

$$|S_n - S| \leq |S_n - S_{N_1,n_2}| + |S_{N_1,n_2} - \bar{S}_{N_1^d}| + |\bar{S}_{N_1^d} - S|.$$

Using the same arguments as before, we show that, P_X-a.s., $|S_n - S| \to 0$ as $n \to +\infty$.

□

Proof of proposition Appendix C.8.

Proof. $\mathbb{E}\left(\frac{1}{n} y^t M y\right) = \mathbb{E}\left\{\mathbb{E}\left(\frac{1}{n} y^t M y | X\right)\right\} = \mathbb{E}\left\{\frac{1}{n}\mathrm{Tr}\left(\mathrm{MR}_0\right)\right\} \to \Sigma$. Furthermore $var\left(\frac{1}{n} y^t M y\right) = \mathbb{E}\left\{var\left(\frac{1}{n} y^t M y | X\right)\right\} + var\left\{\mathbb{E}\left(\frac{1}{n} y^t M y | X\right)\right\}$. $var\left(\frac{1}{n} y^t M y | X = x\right)$ is a $O\left(\frac{1}{n}\right)$, uniformly in x, using proposition Appendix C.7 and $||A + B|| \leq ||A|| + ||B||$. Therefore $var\left(\frac{1}{n} y^t M y | X\right)$ is bounded by $O(\frac{1}{n})$ P_X-a.s. $var\left\{\mathbb{E}\left(\frac{1}{n} y^t M y | X\right)\right\} = var\left\{\frac{1}{n}\mathrm{Tr}\left(\mathrm{MR}_{\theta_0}\right)\right\} \to 0$, using proposition Appendix C.7. Hence $\frac{1}{n} y^t M y$ converges to Σ in the mean square sense.

□

Proof of proposition Appendix C.9.

Proof. Let $v_\lambda = (\lambda_1, ..., \lambda_p) \in \mathbb{R}^p$.

$$\mathbb{E}\left(\exp\left[\mathrm{i}\sum_{k=1}^{p} \lambda_k \frac{1}{\sqrt{n}}\left\{\mathrm{Tr}\left(\mathrm{M_k}\right) + \mathrm{y^t N_k y}\right\}\right]\right) = \mathbb{E}\left\{\mathbb{E}\left(\exp\left[\mathrm{i}\sum_{k=1}^{p} \lambda_k \frac{1}{\sqrt{n}}\left\{\mathrm{Tr}\left(\mathrm{M_k}\right) + \mathrm{y^t N_k y}\right\}\right] | X\right)\right\}.$$

For fixed $x = (x_1, ..., x_n) \in (S_X)^n$, denoting $\sum_{k=1}^p \lambda_k R^{\frac{1}{2}} N_k R^{\frac{1}{2}} = P^t D P$, with $P^t P = I_n$ and D diagonal, $z = P R^{-\frac{1}{2}} y$ (which is a vector of *iid* standard Gaussian variables, conditionally to $X = x$), we have

$$\begin{aligned}\sum_{k=1}^{p} \lambda_k \frac{1}{\sqrt{n}}\left\{\mathrm{Tr}\left(\mathrm{M_k}\right) + \mathrm{y^t N_k y}\right\} &= \frac{1}{\sqrt{n}}\left[\mathrm{Tr}\left(\sum_{\mathrm{k}=1}^{\mathrm{p}} \lambda_{\mathrm{k}} \mathrm{M_k}\right) + \sum_{\mathrm{i}=1}^{\mathrm{n}} \phi_{\mathrm{i}}\left(\sum_{\mathrm{k}=1}^{\mathrm{p}} \lambda_{\mathrm{k}} \mathrm{R}^{\frac{1}{2}} \mathrm{N_k R}^{\frac{1}{2}}\right) \mathrm{z_i^2}\right] \\ &= \frac{1}{\sqrt{n}}\left[\sum_{i=1}^{n} \phi_i\left(\sum_{k=1}^{p} \lambda_k R^{\frac{1}{2}} N_k R^{\frac{1}{2}}\right)\left\{z_i^2 - 1\right\}\right].\end{aligned}$$

Hence

$$\begin{aligned} var\left[\sum_{k=1}^{p}\lambda_k\frac{1}{\sqrt{n}}\left\{\mathrm{Tr}\left(\mathrm{M_k}\right)+\mathrm{y^t N_k y}\right\}\Big|X\right] &= \frac{2}{n}\sum_{i=1}^{n}\phi_i^2\left(\sum_{k=1}^{p}\lambda_k R^{\frac{1}{2}}N_k R^{\frac{1}{2}}\right) \\ &= \frac{2}{n}\sum_{k=1}^{p}\sum_{l=1}^{p}\lambda_k\lambda_l\mathrm{Tr}\left(\mathrm{RN_k RN_l}\right) \\ &\underset{n\to+\infty}{\to} v_\lambda^t\left(2\Sigma\right)v_\lambda \quad \text{for a.e. } \omega_X\ . \end{aligned}$$

Hence, for almost every ω_X, we can apply Lindeberg-Feller criterion to the Ω_Y-measurable variables $\frac{1}{\sqrt{n}}\phi_i\left(\sum_{k=1}^{p}\lambda_k R^{\frac{1}{2}}N_k R^{\frac{1}{2}}\right)\left\{(z_x)_i^2-1\right\}$, $1\le i\le n$, to show that $\sum_{k=1}^{p}\lambda_k\frac{1}{\sqrt{n}}\left\{\mathrm{Tr}\left(\mathrm{M_k}\right)+\mathrm{y^t N_k y}\right\}$ converges in law to $\mathcal{N}\left(0,v_\lambda^t\left(2\Sigma\right)v_\lambda\right)$. Hence, $\mathbb{E}\left(\exp\left[\mathrm{i}\sum_{k=1}^{p}\lambda_k\frac{1}{\sqrt{n}}\left\{\mathrm{Tr}\left(\mathrm{M_k}\right)+\mathrm{y^t N_k y}\right\}\right]\Big|X\right)$ converges for almost every ω_X to $\exp\left(-\frac{1}{2}v_\lambda^t\left(2\Sigma\right)v_\lambda\right)$. Using the dominated convergence theorem on $(\Omega_X,\mathcal{F}_X,P_X)$, $\mathbb{E}\left(\exp\left[\mathrm{i}\sum_{k=1}^{p}\lambda_k\frac{1}{\sqrt{n}}\left\{\mathrm{Tr}\left(\mathrm{M_k}\right)+\mathrm{y^t N_k y}\right\}\right]\right)$ converges to $\exp\left\{-\frac{1}{2}v_\lambda^t\left(2\Sigma\right)v_\lambda\right\}$. □

Proof of proposition Appendix C.10.

Proof. It is enough to consider the case $c\left(\hat{\theta}\right)=0$, the case $P\left\{c\left(\hat{\theta}\right)=0\right\}\to 1$ being deduced from it by modifying c on a set with vanishing probability measure, which does not affect the convergence in law. For all $1\le k\le p$

$$0=c_k\left(\hat{\theta}\right)=c_k\left(\theta_0\right)+\left\{\frac{\partial}{\partial\theta}c_k\left(\theta_0\right)\right\}^t\left(\hat{\theta}-\theta_0\right)+r,$$

with random r, so that $|r|\le\sup_{\tilde{\theta},i,j,k}\left|\frac{\partial^2}{\partial\theta_i\partial\theta_j}c_k\left(\tilde{\theta}\right)\right|\times|\hat{\theta}-\theta_0|^2$. Hence $r=o_p\left(|\hat{\theta}-\theta_0|\right)$. We then have

$$-c_k\left(\theta_0\right) = \left[\left\{\frac{\partial}{\partial\theta}c_k\left(\theta_0\right)\right\}^t+o_p\left(1\right)\right]\left(\hat{\theta}-\theta_0\right),$$

and so

$$\left(\hat{\theta}-\theta_0\right)=-\left\{\frac{\partial}{\partial\theta}c\left(\theta_0\right)+o_p\left(1\right)\right\}^{-1}c\left(\theta_0\right). \tag{C.2}$$

We conclude using Slutsky lemma.

Remark. One can show that, with probability going to one as $n\to+\infty$, the likelihood has a unique global minimizer. Indeed, we first notice that the set of the minimizers is a subset of any open ball of center θ_0 with probability going to one. For a small enough open ball, the probability that the likelihood function is strictly convex on this open ball converges to one. This is because of the third-order regularity of the likelihood with respect to θ, and because the limit of the second derivative matrix of the Likelihood at θ_0 is positive.

□

Appendix D. Exact expressions of the asymptotic variances at $\epsilon=0$ for $d=1$

In this section we only address the case $d=1$ and $p=1$, where the observation points $v_i+\epsilon X_i$, $1\le i\le n$, $n\in\mathbb{N}^*$, are the $i+\epsilon X_i$, where X_i is uniform on $[-1,1]$, and $\Theta=[\theta_{inf},\theta_{sup}]$.

We define the Fourier transform function $\hat{s}\left(.\right)$ of a sequence s_n of $\mathbb{Z}$ by $\hat{s}\left(f\right)=\sum_{n\in\mathbb{Z}}s_n e^{\mathrm{i}s_n f}$ as in [8]. This function is 2π periodic on $[-\pi,\pi]$. Then

- The sequence of the $K_{\theta_0}\left(i\right)$, $i\in\mathbb{Z}$, has Fourier transform f which is even and non-negative on $[-\pi,\pi]$.

- The sequence of the $\frac{\partial}{\partial\theta}K_{\theta_0}(i)$, $i\in\mathbb{Z}$, has Fourier transform f_θ which is even on $[-\pi,\pi]$.
- The sequence of the $\frac{\partial}{\partial t}K_{\theta_0}(i)\mathbf{1}_{i\neq 0}$, $i\in\mathbb{Z}$, has Fourier transform i f_t which is odd and imaginary on $[-\pi,\pi]$.
- The sequence of the $\frac{\partial}{\partial t}\frac{\partial}{\partial\theta}K_{\theta_0}(i)\mathbf{1}_{i\neq 0}$, $i\in\mathbb{Z}$, has Fourier transform i $f_{t,\theta}$ which is odd and imaginary on $[-\pi,\pi]$.
- The sequence of the $\frac{\partial^2}{\partial t^2}K_{\theta_0}(i)\mathbf{1}_{i\neq 0}$, $i\in\mathbb{Z}$, has Fourier transform $f_{t,t}$ which is even on $[-\pi,\pi]$.
- The sequence of the $\frac{\partial^2}{\partial t^2}\frac{\partial}{\partial\theta}K_{\theta_0}(i)\mathbf{1}_{i\neq 0}$, $i\in\mathbb{Z}$, has Fourier transform $f_{t,t,\theta}$ which is even on $[-\pi,\pi]$.

In this section we assume in condition Appendix D.1 that all these sequences are dominated by a decreasing exponential function, so that the Fourier transforms are C^∞. This condition could be weakened, but it simplifies the proofs, and it is satisfied in our framework.

Condition Appendix D.1. *There exist $C<\infty$ and $a>0$ so that the sequences of general terms $K_{\theta_0}(i)$, $\frac{\partial}{\partial\theta}K_{\theta_0}(i)$, $\frac{\partial}{\partial t}K_{\theta_0}(i)\mathbf{1}_{i\neq 0}$, $\frac{\partial}{\partial t}\frac{\partial}{\partial\theta}K_{\theta_0}(i)\mathbf{1}_{i\neq 0}$, $\frac{\partial^2}{\partial t^2}K_{\theta_0}(i)\mathbf{1}_{i\neq 0}$, $\frac{\partial^2}{\partial t^2}\frac{\partial}{\partial\theta}K_{\theta_0}(i)\mathbf{1}_{i\neq 0}$, $i\in\mathbb{Z}$, are bounded by $Ce^{-a|i|}$.*

For a 2π-periodic function f on $[-\pi,\pi]$, we denote by $M(f)$ the mean value of f on $[-\pi,\pi]$.

Then, proposition Appendix D.2 gives the closed form expressions of Σ_{ML}, $\Sigma_{CV,1}$, $\Sigma_{CV,2}$ and $\frac{\partial^2}{\partial\epsilon^2}\Sigma_{ML}\Big|_{\epsilon=0}$.

Proposition Appendix D.2. *Assume that conditions 2.1 and Appendix D.1 are verified.*

For $\epsilon=0$,

$$\Sigma_{ML}=M\left(\frac{f_\theta^2}{f^2}\right),$$

$$\begin{aligned}\Sigma_{CV,1} &= 8M\left(\frac{1}{f}\right)^{-6}M\left(\frac{f_\theta}{f^2}\right)^2M\left(\frac{1}{f^2}\right)\\ &\quad +8M\left(\frac{1}{f}\right)^{-4}M\left(\frac{f_\theta^2}{f^4}\right)\\ &\quad -16M\left(\frac{1}{f}\right)^{-5}M\left(\frac{f_\theta}{f^2}\right)M\left(\frac{f_\theta}{f^3}\right),\end{aligned}$$

$$\Sigma_{CV,2} = 2M\left(\frac{1}{f}\right)^{-3}\left\{M\left(\frac{f_\theta^2}{f^3}\right)M\left(\frac{1}{f}\right)-M\left(\frac{f_\theta}{f^2}\right)^2\right\},$$

and

$$\begin{aligned}\left.\frac{\partial^2}{\partial\epsilon^2}\Sigma_{ML}\right|_{\epsilon=0} &= \frac{4}{3}M\left(\frac{f_\theta}{f^2}\right)M\left(\frac{f_t^2\ f_\theta}{f^2}\right)\\ &- \frac{8}{3}M\left(\frac{1}{f}\right)M\left(\frac{f_{t,\theta}\ f_t\ f_\theta}{f^2}\right)-\frac{8}{3}M\left(\frac{f_\theta}{f^2}\right)M\left(\frac{f_{t,\theta}\ f_t}{f}\right)\\ &+ \frac{4}{3}M\left(\frac{1}{f}\right)M\left(\frac{f_t^2\ f_\theta^2}{f^3}\right)+\frac{4}{3}M\left(\frac{f_\theta^2}{f^3}\right)M\left(\frac{f_t^2}{f}\right)\\ &- \frac{4}{3}M\left(\frac{f_{t,t}\ f_\theta^2}{f^3}\right)\\ &+ \frac{4}{3}M\left(\frac{1}{f}\right)M\left(\frac{f_{t,\theta}^2}{f}\right)\\ &+ \frac{4}{3}M\left(\frac{f_{t,t,\theta}\ f_\theta}{f^2}\right).\end{aligned}$$

Proposition Appendix D.2 is proved in the supplementary material.

An interesting remark can be made on $\Sigma_{CV,2}$. Using Cauchy-Schwartz inequality, we obtain

$$\Sigma_{CV,2}=2M\left(\frac{1}{f}\right)^{-3}\left[M\left\{\left(\frac{f_\theta}{f^{\frac{3}{2}}}\right)^2\right\}M\left\{\left(\frac{1}{f^{\frac{1}{2}}}\right)^2\right\}-M\left\{\frac{f_\theta}{f^{\frac{3}{2}}}\frac{1}{f^{\frac{1}{2}}}\right\}^2\right]\geq 0,$$

so that the limit of the second derivative with respect to θ of the CV criterion at θ_0 is indeed non-negative. Furthermore, for the limit to be zero, it is necessary that $\frac{f_\theta}{f^{\frac{3}{2}}}$ be proportional to $\frac{1}{f^{\frac{1}{2}}}$, that is to say f_θ be proportional to f. This is equivalent to $\frac{\partial K_{\theta_0}}{\partial\theta}$ being proportional to K_{θ_0} on $\mathbb{Z}$, which happens only when around θ_0, $K_\theta(i)=\frac{\theta}{\theta_0}K_{\theta_0}(i)$, for $i\in\mathbb{Z}$. Hence around θ_0, θ would be a global variance hyper-parameter. Therefore, we have shown that for the regular grid in dimension one, the asymptotic variance is positive, as long as θ is not only a global variance hyper-parameter.

References

[1] P. Abrahamsen, A review of Gaussian random fields and correlation functions, Technical Report, Norwegian computing center, 1997.

[2] E. Anderes, On the consistent separation of scale and variance for Gaussian random fields, The Annals of Statistics 38 (2010) 870–893.

[3] F. Bachoc, Cross validation and maximum likelihood estimations of hyper-parameters of Gaussian processes with model mispecification (+2012). In revision for Computational Statistics and Data Analysis.

[4] E.J. Candes, T. Tao, Near-optimal signal recovery from random projections: Universal encoding strategies?, Information Theory, IEEE Transactions on 52 (2006) 5406 –5425.

[5] N. Cressie, S. Lahiri, The asymptotic distribution of reml estimators, Journal of Multivariate Analysis 45 (1993) 217–233.

[6] J. Du, H. Zhang, V. Mandrekar, Fixed domain asymptotics properties of tapered maximum likelihood estimators, The Annals of Statistics 37 (2009) 3330–3361.

[7] O. Dubrule, Cross validation of Kriging in a unique neighborhood, Mathematical Geology 15 (1983) 687–699.

[8] R. Gray, Toeplitz and Circulant Matrices: A review, Technical Report, 2001.

[9] D. Jones, M. Schonlau, W. Welch, Efficient global optimization of expensive black box functions, Journal of Global Optimization 13 (1998) 455–492.

[10] W. Loh, Fixed domain asymptotics for a subclass of Matérn type Gaussian random fields, The Annals of Statistics 33 (2005) 2344–2394.

[11] W. Loh, T. Lam, Estimating structured correlation matrices in smooth Gaussian random field models, The Annals of Statistics 28 (2000) 880–904.

[12] K. Mardia, R. Marshall, Maximum likelihood estimation of models for residual covariance in spatial regression, Biometrika 71 (1984) 135–146.

[13] D.C. Montgomery, Design and Analysis of Experiments, Wiley, New York, 2005. 6th edition.

[14] H. Niederreiter, Random Number Generation and Quasi-Monte Carlo Methods, series SIAM CBMS-NSF, SIAM, Philadelphia, 1992.

[15] R. Paulo, G. Garcia-Donato, J. Palomo, Calibration of computer models with multivariate output, Computational Statistics and Data Analysis 56 (2012) 3959–3974.

[16] W.H. Press, S.A. Teukolsky, W. Vetterling, B. Flannery, Numerical recipes: The art of Scientific computing, Cambridge university press, 2007. 3rd edition.

[17] C. Rasmussen, C. Williams, Gaussian Processes for Machine Learning, The MIT Press, Cambridge, 2006.

[18] B. Ripley, Spatial Statistics, Wiley, New York, 1981.

[19] J. Sacks, W. Welch, T. Mitchell, H. Wynn, Design and analysis of computer experiments, Statistical Science 4 (1989) 409–423.

[20] T. Santner, B. Williams, W. Notz, The Design and Analysis of Computer Experiments, Springer, New York, 2003.

[21] M. Stein, Asymptotically efficient prediction of a random field with a misspecified covariance function, The Annals of Statistics 16 (1988) 55–63.

[22] M. Stein, Bounds on the efficiency of linear predictions using an incorrect covariance function, The Annals of Statistics 18 (1990) 1116–1138.

[23] M. Stein, Uniform asymptotic optimality of linear predictions of a random field using an incorrect second-order structure, The Annals of Statistics 18 (1990) 850–872.

[24] M. Stein, Interpolation of Spatial Data: Some Theory for Kriging, Springer, New York, 1999.

[25] S. Sundararajan, S. Keerthi, Predictive approaches for choosing hyperparameters in Gaussian processes, Neural Computation 13 (2001) 1103–1118.

[26] T. Sweeting, Uniform asymptotic normality of the maximum likelihood estimator, The Annals of Statistics 8 (1980) 1375–1381.

[27] E. Vazquez, Modélisation comportementale de systèmes non-linéaires multivariables par méthodes à noyaux et applications, Ph.D. thesis, Université Paris XI Orsay, 2005. Available at http://tel.archives-ouvertes.fr/tel-00010199/en.

[28] Z. Ying, Asymptotic properties of a maximum likelihood estimator with data from a Gaussian process, Journal of Multivariate Analysis 36 (1991) 280–296.

[29] Z. Ying, Maximum likelihood estimation of parameters under a spatial sampling scheme, The Annals of Statistics 21 (1993) 1567–1590.

[30] H. Zhang, Inconsistent estimation and asymptotically equivalent interpolations in model-based geostatistics, Journal of the American Statistical Association 99 (2004) 250–261.

[31] H. Zhang, Y. Wang, Kriging and cross validation for massive spatial data, Environmetrics 21 (2010) 290–304.

[32] Z. Zhu, H. Zhang, Spatial sampling design under the infill asymptotics framework, Environmetrics 17 (2006) 323–337.